

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2003, 14:09:58 ; Search time 69 Seconds
(without alignments)
21.243 Million cell updates/sec

Title: US-09-909-164-42
Perfect score: 52
Sequence: 1 EEVVPXGXHS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*			
1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*		
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*		
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*		
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*		
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*		
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*		
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*		
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*		
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*		
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*		
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*		
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*		
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*		
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*		
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*		
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*		
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*		
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*		
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*		
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*		
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*		
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*		
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	49	94.2	11	23	ABB80546 Hepatitis C virus
2	49	94.2	11	23	ABB80550 Hepatitis C virus
3	49	94.2	11	23	ABB80554 Hepatitis C virus
4	49	94.2	11	23	ABB80555 Hepatitis C virus
5	48	92.3	11	23	ABB80523 Hepatitis C virus
6	48	92.3	11	23	ABB80527 Hepatitis C virus
7	48	92.3	11	23	ABB80531 Hepatitis C virus
8	48	92.3	11	23	ABB80532 Hepatitis C virus
9	48	92.3	11	23	ABB80537 Hepatitis C virus
10	48	92.3	11	23	ABB80541 Hepatitis C virus

11	48	92.3	11	23	ABB80558 Hepatitis C virus
12	48	92.3	11	23	ABB80560 Hepatitis C virus
13	40	76.9	11	23	ABB80544 Hepatitis C virus
14	40	76.9	11	23	ABB80545 Hepatitis C virus
15	40	76.9	11	23	ABB80547 Hepatitis C virus
16	40	76.9	11	23	ABB80548 Hepatitis C virus
17	40	76.9	11	23	ABB80549 Hepatitis C virus
18	40	76.9	11	23	ABB80551 Hepatitis C virus
19	40	76.9	11	23	ABB80552 Hepatitis C virus
20	40	76.9	11	23	ABB80553 Hepatitis C virus
21	40	76.9	11	23	ABB80556 Hepatitis C virus
22	40	76.9	11	23	ABB80557 Hepatitis C virus
23	39	75.0	11	23	ABB80521 Hepatitis C virus
24	39	75.0	11	23	ABB80522 Hepatitis C virus
25	39	75.0	11	23	ABB80524 Hepatitis C virus
26	39	75.0	11	23	ABB80525 Hepatitis C virus
27	39	75.0	11	23	ABB80526 Hepatitis C virus
28	39	75.0	11	23	ABB80528 Hepatitis C virus
29	39	75.0	11	23	ABB80529 Hepatitis C virus
30	39	75.0	11	23	ABB80530 Hepatitis C virus
31	39	75.0	11	23	ABB80533 Hepatitis C virus
32	39	75.0	11	23	ABB80534 Hepatitis C virus
33	39	75.0	11	23	ABB80535 Hepatitis C virus
34	39	75.0	11	23	ABB80536 Hepatitis C virus
35	39	75.0	11	23	ABB80538 Hepatitis C virus
36	39	75.0	11	23	ABB80539 Hepatitis C virus
37	39	75.0	11	23	ABB80540 Hepatitis C virus
38	39	75.0	11	23	ABB80542 Hepatitis C virus
39	39	75.0	11	23	ABB80543 Hepatitis C virus
40	39	75.0	11	23	ABB80559 Hepatitis C virus
41	39	75.0	11	23	ABB80561 Hepatitis C virus
42	39	75.0	11	23	ABB80562 Hepatitis C virus
43	39	75.0	11	23	ABB80563 Hepatitis C virus
44	39	75.0	11	23	ABB80564 Hepatitis C virus
45	39	75.0	11	23	ABB80565 Hepatitis C virus

ALIGNMENTS

RESULT 1	
ABB80546	
ID	ABB80546 standard; peptide; 11 AA.
XX	
AC	ABB80546;
XX	
DT	08-OCT-2002 (first entry)
XX	
DE	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #26.
XX	
KW	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.
XX	
OS	Synthetic.
XX	
PH	Key
FT	Location/Qualifiers
FT	Modified-site 1
FT	/note= "N-terminal acetyl"
FT	Modified-site 6
FT	/note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
FT	Modified-site 11
FT	/note= "C-terminal amide"
XX	
PN	WO200208251-A2.
XX	
PD	31-JAN-2002.
XX	
PF	19-JUL-2001; 2001WO-US23169.
XX	
PR	21-JUL-2000; 2000US-220101P.
XX	
PA	(CORV-) CORVAS INT INC.

XX Lim-wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX Sequence 11 AA;

Query Match 94.2%; Score 49; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.0034;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYS 11
 DB 1 EEVVPXGXHYS 11

RESULT 2
 ABB80550
 ID ABB80550 standard; peptide; 11 AA.
 XX AC ABB80550;
 XX 08-OCT-2002 (first entry)
 XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #30.
 DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 KW Synthetic.

XX Key Location/Qualifiers
 FH Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT WO200208251-A2.
 XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US23169.
 XX 21-JUL-2000; 2000US-220101P.
 XX (CORV-) CORVAS INT INC.

XX Lim-wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -

XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX Sequence 11 AA;

Query Match 94.2%; Score 49; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.0034;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYS 11
 DB 1 EEVVPXGXHYS 11

RESULT 3
 ABB80554
 ID ABB80554 standard; peptide; 11 AA.

XX AC ABB80554;
 XX 08-OCT-2002 (first entry)

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #34.
 XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 KW Synthetic.

XX Key Location/Qualifiers
 FH Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT WO200208251-A2.
 XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US23169.
 XX 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.
 XX Lim-wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -

XX Claim 17; Page 65; 69pp; English.

XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A

CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;
 Query Match 94.2%; Score 49; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.0034;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EEVVPXGXHYS 11
 DB 1 EEVVPXGSHYS 11

RESULT 4

ABB80555
 ID ABB80555 standard; peptide; 11 AA.

XX
 AC ABB80555;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #35.

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.

OS Synthetic.

Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

FT WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US23169.

XX 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
 activity useful for treating disorders associated with hepatitis C
 virus protease

XX Claim 17; Page 65; 69pp; English.

XX The sequence represents a peptide compound of the invention having
 hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 invention are alpha-ketoamide peptide analogues. The peptides have
 virucide activity, and are useful for treating and in the manufacture of
 a medicament to treat disorders associated with HCV protease. A
 pharmaceutical composition comprising the peptide as an active ingredient
 is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 94.2%; Score 49; DB 23; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.0034;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EEVVPXGXHYS 11
 DB 1 EEVVPXGSHYS 11

RESULT 5

ABB80523
 ID ABB80523 standard; peptide; 11 AA.

XX
 AC ABB80523;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #3.

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.

OS Synthetic.

Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

FT WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US23169.

XX 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
 activity useful for treating disorders associated with hepatitis C
 virus protease

XX Claim 17; Page 64; 69pp; English.

XX The sequence represents a peptide compound of the invention having
 hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 invention are alpha-ketoamide peptide analogues. The peptides have
 virucide activity, and are useful for treating and in the manufacture of
 a medicament to treat disorders associated with HCV protease. A
 pharmaceutical composition comprising the peptide as an active ingredient
 is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 92.3%; Score 48; DB 23; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.0054;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EEVVPXGXHYS 11
 DB 1 EEVVPXGMHYS 11

```

RESULT 6
ABB80527
ID ABB80527 standard; peptide; 11 AA.
XX
AC ABB80527;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #7.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX
OS Synthetic.
XX
FH Key
FT Modified-site 1 Location/Qualifiers
FT /note= "N-terminal acetyl"
FT Modified-site 6
FT /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 11
FT /note= "C-terminal amide"
XX
PN WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-wilby M, Levy OE, Brunck TK;
XX
PD WPI; 2002-361643/39.
XX
PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease
XX
PS Claim 17; Page 64; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Query Match 92.3%; Score 48; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0054;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYS 11
DB 1 EEVVPXGGHYS 11

RESULT 7
ABB80531
ID ABB80531 standard; peptide; 11 AA.
XX
AC ABB80531;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #12.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX
OS Synthetic.
XX
FH Key
FT Modified-site 1 Location/Qualifiers
FT /note= "N-terminal acetyl"
FT Modified-site 6
FT /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 11
FT /note= "C-terminal amide"
XX
PN WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-wilby M, Levy OE, Brunck TK;
XX
PD WPI; 2002-361643/39.
XX
PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease
XX
PS Claim 17; Page 64; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Query Match 92.3%; Score 48; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0054;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYS 11
DB 1 EEVVPXGGHYS 11

RESULT 8
ABB80532
ID ABB80532 standard; peptide; 11 AA.
XX
AC ABB80532;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #12.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX
OS Synthetic.
XX
FH Key
FT Modified-site 1 Location/Qualifiers
FT /note= "N-terminal acetyl"

```

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT FT 11
FT Modified-site 11 /note= "C-terminal amide"
FT FT
FT WO200208251-A2.
FT PN
FT PD 31-JAN-2002.
FT PF 19-JUL-2001; 2001WO-US23169.
FT XX 21-JUL-2000; 2000US-220101P.
FT PR (CORV-) CORVAS INT INC.
FT XX
FT PA Lim-wilby M, Levy OE, Brunck TK;
FT PI WPI; 2002-361643/39.
FT DR Novel peptide compound having hepatitis C virus protease inhibitory
FT PT activity useful for treating disorders associated with hepatitis C
FT PT virus protease
FT XX
FT PS Claim 17; Page 64; 69pp; English.
FT XX
FT CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
FT XX
SQ Sequence 11 AA;
Query Match 92.3%; Score 48; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0054;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGXHYS 11
DB 1 EEVVPXGXHYS 11
RESULT 9
ABB80537
ID ABB80537 standard; peptide; 11 AA.
XX AC ABB80537;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #17.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX KW virucide.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT FT residue 7"
FT FT Misc-difference 9 /note= "D-form residue"
FT FT Modified-site 11 /note= "C-terminal amide"
FT FT
FT PN WO200208251-A2.
FT XX 31-JAN-2002.
FT PF 19-JUL-2001; 2001WO-US23169.
FT XX 21-JUL-2000; 2000US-220101P.
FT PR (CORV-) CORVAS INT INC.
FT XX
FT PA Lim-wilby M, Levy OE, Brunck TK;
FT PI WPI; 2002-361643/39.
FT DR Novel peptide compound having hepatitis C virus protease inhibitory
FT PT activity useful for treating disorders associated with hepatitis C
FT PT virus protease
FT XX
FT PS Claim 17; Page 64; 69pp; English.
FT XX
FT CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
FT XX
SQ Sequence 11 AA;
Query Match 92.3%; Score 48; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0054;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGXHYS 11
DB 1 EEVVPXGXHYS 11
RESULT 10
ABB80541
ID ABB80541 standard; peptide; 11 AA.
XX AC ABB80541;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #21.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX KW virucide.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT FT residue 7"
FT FT Misc-difference 8 /note= "D-form residue"
FT FT Modified-site 11 /note= "C-terminal amide"
FT FT
FT PN WO200208251-A2.
FT XX 31-JAN-2002.
FT PF 19-JUL-2001; 2001WO-US23169.
FT XX 21-JUL-2000; 2000US-220101P.
FT PR
FT XX

PN WO200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US23169.
XX 21-JUL-2000; 2000US-220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C
XX virus protease
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 92.3%; Score 48; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0054;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGXHYS 11
DB 1 EEVVPXGXHYS 11
RESULT 10
ABB80541
ID ABB80541 standard; peptide; 11 AA.
XX AC ABB80541;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #21.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX KW virucide.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT FT residue 7"
FT FT Misc-difference 8 /note= "D-form residue"
FT FT Modified-site 11 /note= "C-terminal amide"
FT FT
FT PN WO200208251-A2.
FT XX 31-JAN-2002.
FT PF 19-JUL-2001; 2001WO-US23169.
FT XX 21-JUL-2000; 2000US-220101P.
FT PR
FT XX

PA (CORV-) CORVAS INT INC.
 XX Lim-wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX Sequence 11 AA;
 SQ

Query Match 92.3%; Score 48; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.0054;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYS 11
 ||||| ||||
 Db 1 EEVVPXGXHYS 11

RESULT 11
 ABB80558
 ID ABB80558 standard; peptide; 11 AA.
 XX
 AC ABB80558;
 XX
 DT 08-OCT-2002 (first entry)
 XX
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #38.
 XX
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Modified-site 8 /note= "Oxymethionine"
 FT Modified-site 11 /note= "C-terminal amide"
 FT
 FT WO200208251-A2.
 PN
 XX 31-JAN-2002.
 XX
 XX 19-JUL-2001; 2001WO-US23169.
 XX
 XX 21-JUL-2000; 2000US-220101P.
 XX
 XX (CORV-) CORVAS INT INC.
 XX
 XX Lim-wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C

PT virus protease -
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX Sequence 11 AA;
 SQ

Query Match 92.3%; Score 48; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.0054;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYS 11
 ||||| ||||
 Db 1 EEVVPXGXHYS 11

RESULT 12
 ABB80560
 ID ABB80560 standard; peptide; 11 AA.
 XX
 AC ABB80560;
 XX
 DT 08-OCT-2002 (first entry)
 XX
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #40.
 XX
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8 /note= "D-form residue"
 FT Modified-site 8 /note= "Oxymethionine"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT
 FT WO200208251-A2.
 PN
 XX 31-JAN-2002.
 XX
 XX 19-JUL-2001; 2001WO-US23169.
 XX
 XX 21-JUL-2000; 2000US-220101P.
 XX
 XX (CORV-) CORVAS INT INC.
 XX
 XX Lim-wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -
 XX Claim 17; Page 65; 69pp; English.

CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 92.3%; Score 48; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0054;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYHS 11
|||||
Db 1 EEVVPXGXHYHS 11

RESULT 13

ABB80544
ID ABB80544 standard; peptide; 11 AA.

XX AC ABB80544;

XX DT 08-OCT-2002 (first entry)

XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #24.

XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Modified-site 1

FT /note= "N-terminal acetyl"

FT Modified-site 6

FT /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"

FT Modified-site 11

FT /note= "C-terminal amide"

FT WO200208251-A2.

XX PD 31-JAN-2002.

XX PF 19-JUL-2001; 2001WO-US23169.

XX PR 21-JUL-2000; 2000US-220101P.

XX PA (CORV-) CORVAS INT INC.

XX PI Lim-wilby M, Levy OE; Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease

XX Claim 17; Page 65; 69pp; English.

CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 76.9%; Score 40; DB 23; Length 11;
Best Local Similarity 81.8%; Pred. No. 0.21;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYHS 11
|||||
Db 1 EEVVPXGXHYHS 11

RESULT 14

ABB80545
ID ABB80545 standard; peptide; 11 AA.

XX AC ABB80545;

XX DT 08-OCT-2002 (first entry)

XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #25.

XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Modified-site 1

FT /note= "N-terminal acetyl"

FT Modified-site 6

FT /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"

FT Misc-difference 9

FT /note= "D-form residue"

FT Modified-site 11

FT /note= "C-terminal amide"

FT WO200208251-A2.

XX PD 31-JAN-2002.

XX PF 19-JUL-2001; 2001WO-US23169.

XX PR 21-JUL-2000; 2000US-220101P.

XX PA (CORV-) CORVAS INT INC.

XX PI Lim-wilby M, Levy OE; Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease

XX Claim 17; Page 65; 69pp; English.

CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 76.9%; Score 40; DB 23; Length 11;
Best Local Similarity 81.8%; Pred. No. 0.21;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYHS 11
|||||
Db 1 EEVVPXGXHYHS 11

RESULT 15
ABB80547
ID ABB80547 standard; peptide; 11 AA.
XX
AC ABB80547;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #27.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6
FT /note= "Norvalyl carbonyl forming keto-amide linkage with
residue 7"
FT Modified-site 11
FT /note= "C-terminal amide"
XX
PN WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-wilby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
PT Novel peptide compound having hepatitis C virus protease inhibitory
activity useful for treating disorders associated with hepatitis C
virus protease
XX
PS Claim 17; Page 65; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
invention are alpha-ketoamide peptide analogues. The peptides have
virucide activity, and are useful for treating and in the manufacture of
a medicament to treat disorders associated with HCV protease. A
pharmaceutical composition comprising the peptide as an active ingredient
is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 76.9%; Score 40; DB 23; Length 11;
Best Local Similarity 81.8%; Pred. No. 0.21;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 EEVVPXGXHYS 11
|||||||
Db 1 EEVVPXGTDYS 11
Search completed: June 13, 2003, 14:11:15
Job time : 69 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2003, 14:10:03 ; Search time 26 Seconds
(without alignments)
12.448 Million cell updates/sec

Title: US-09-909-164-42
Perfect score: 52
Sequence: 1 EEVVPXGXHYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574.

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2.6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2.6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2.6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2.6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2.6/ptodata/1/1aa/PCTUS_COMB.pep.*
6: /cgn2.6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	36	69.2	1037	4	US-09-134-001C-4794
2	34	65.4	126	2	US-08-879-995A-3
3	34	65.4	126	3	US-09-215-096-3
4	33	63.5	447	4	US-08-961-083-182
5	33	63.5	600	2	US-08-821-119-19
6	33	63.5	600	2	US-08-821-118-2
7	33	63.5	763	4	US-08-961-083-66
8	33	63.5	796	4	US-08-961-083-56
9	32	61.5	738	1	US-08-530-010-3
10	32	61.5	738	1	US-08-530-010-5
11	32	61.5	738	1	US-08-530-010-7
12	32	61.5	738	1	US-08-530-010-9
13	32	61.5	738	1	US-08-530-010-11
14	32	61.5	738	2	US-08-484-101B-3
15	32	61.5	738	2	US-08-484-101B-5
16	32	61.5	738	2	US-08-484-101B-7
17	32	61.5	738	2	US-08-484-101B-9
18	32	61.5	738	2	US-08-484-101B-11
19	32	61.5	738	4	US-08-714-524D-5
20	32	61.5	738	4	US-08-714-524D-7
21	32	61.5	738	4	US-08-714-524D-9
22	32	61.5	738	4	US-08-714-524D-11
23	32	61.5	738	4	US-08-714-524D-13
24	31	59.6	70	4	US-09-134-001C-3950
25	31	59.6	237	4	US-09-134-001C-4185
26	31	59.6	622	4	US-09-342-647-4
27	31	59.6	602	2	US-08-459-146-2

28	31	59.6	622	2	US-08-459-065-2	Sequence 2, Appli
29	31	59.6	630	4	US-09-342-647-2	Sequence 2, Appli
30	31	59.6	667	4	US-09-342-647-28	Sequence 8, Appli
31	30.5	58.7	268	4	US-09-085-305-8	Sequence 10, Appli
32	30.5	58.7	268	4	US-09-085-305-10	Sequence 17, Appli
33	30.5	58.7	268	4	US-09-085-305-17	Sequence 68, Appli
34	30	57.7	121	4	US-09-152-060-68	Sequence 85, Appli
35	30	57.7	121	4	US-09-152-060-85	Sequence 1, Appli
36	30	57.7	122	2	US-08-879-995A-1	Sequence 1, Appli
37	30	57.7	122	3	US-09-215-096-1	Sequence 2, Appli
38	30	57.7	241	3	US-08-834-776A-2	Sequence 10, Appli
39	30	57.7	328	4	US-09-180-827-10	Sequence 2, Appli
40	30	57.7	386	4	US-09-086-483A-2	Sequence 6, Appli
41	30	57.7	386	4	US-09-130-491-6	Sequence 17, Appli
42	30	57.7	615	2	US-08-563-566A-17	Sequence 17, Appli
43	30	57.7	615	2	US-08-023-610-17	Sequence 17, Appli
44	30	57.7	615	2	US-08-288-065A-17	Sequence 17, Appli
45	30	57.7	615	2	US-08-362-240A-17	Sequence 17, Appli

ALIGNMENTS

RESULT 1
US-09-134-001C-4794
; Sequence 4794, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4794
; LENGTH: 1037
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4794

Query Match 69.2%; Score 36; DB 4; Length 1037;
Best Local Similarity 63.6%; Pred. No. 37;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EEVVPXGXHYS 11
Db 199 KEVVSNGLHYS 209

RESULT 2
US-08-879-995A-3
; Sequence 3, Application US/08879995A
; Patent No. 5985606
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: HUMAN PREPROTACHYKININ B
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/879,995A
FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0326 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 163590
US-08-879-995A-3

Query Match 65.4%; Score 34; DB 2; Length 126;
Best Local Similarity 66.7%; Pred. No. 9;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGXH 9
Db 28 EQVVPGGGH 36

RESULT 3
US-09-215-096-3
Sequence 3, Application US/09215096
Patent No. 6008194
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Kaser, Matthew R.
TITLE OF INVENTION: HUMAN PREPROTACHYKININ B
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/215,096
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/879,995
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0326 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 163590
US-09-215-096-3

Query Match 65.4%; Score 34; DB 3; Length 126;
Best Local Similarity 66.7%; Pred. No. 9;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGXH 9
Db 28 EQVVPGGGH 36

RESULT 4
US-08-961-083-182
Sequence 182, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi, et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 182:
SEQUENCE CHARACTERISTICS:
LENGTH: 447 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-182

Query Match 63.5%; Score 33; DB 4; Length 447;
Best Local Similarity 62.5%; Pred. No. 58;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVVPXGXHY 10
Db 178 IVPHGHHY 185

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/821,118
; FILING DATE: 19-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4107,400-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 600 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-821-118-2

Query Match          63.5%; Score 33; DB 2: Length 600;
Best Local Similarity 75.0%; Pred.No. 80;
Matches      6; Conservative    0; Mismatches   2; Indels     0; Gaps    0;

QY      4 VPKGXHYS 11
        ||| ||||
Db       31 VPKGWHYS 38

RESULT 7
US-08-961-083-66
; Sequence 66, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids

```

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-66

Query Match 63.5%; Score 33; DB 4; Length 763;
Best Local Similarity 62.5%; Pred. No. 1e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
: || | | |
Db 184 IVPHGDIY 191

RESULT 8

US-08-961-083-56
Sequence 56, Application US/08961083
Patent No. 6159469

GENERAL INFORMATION:

APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083

FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 796 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-56

Query Match 63.5%; Score 33; DB 4; Length 796;
Best Local Similarity 62.5%; Pred. No. 1.1e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
: || | | |
Db 185 IVPHGDIY 192

RESULT 9

US-08-530-010-3
Sequence 3, Application US/08530010
Patent No. 5689055

GENERAL INFORMATION:

APPLICANT: Meyerowitz, Elliott M.

APPLICANT: Chang, Caren
APPLICANT: Blecker, Anthony B.
TITLE OF INVENTION: PLANTS HAVING MODIFIED RESPONSE TO ETHYLENE
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Richard F. Trecartin
STREET: 3400 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/530,010
FILING DATE:
CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/086,555
FILING DATE: 01-JUL-1993

ATTORNEY/AGENT INFORMATION:

NAME: Trecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-57515/RET
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 738 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-530-010-3

Query Match

Best Local Similarity 61.5%; Score 32; DB 1; Length 738;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
: || | | |
Db 498 VPTGSHF 505

RESULT 10

US-08-530-010-5

Sequence 5, Application US/08530010
Patent No. 5689055

GENERAL INFORMATION:

APPLICANT: Meyerowitz, Elliott M.
APPLICANT: Chang, Caren
APPLICANT: Blecker, Anthony B.
TITLE OF INVENTION: PLANTS HAVING MODIFIED RESPONSE TO ETHYLENE
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Richard F. Trecartin
STREET: 3400 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/530,010
FILING DATE:
CLASSIFICATION: 800

; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/086,555
 ; FILING DATE: 01-JUL-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Trecartin, Richard F.
 ; REGISTRATION NUMBER: 31,801
 ; REFERENCE/DOCKET NUMBER: A-57515/RFT
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 781-1989
 ; TELEFAX: (415) 398-3249
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 738 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-530-010-5

Query Match 61.5%; Score 32; DB 1; Length 738;
 Best Local Similarity 62.5%; Pred. No. 1.6e+02;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXY 10
 ||| | |
 Db 498 VVPTGSFH 505

RESULT 11
 US-08-530-010-7
 ; Sequence 7, Application US/08530010
 ; Patent No. 5689055
 ; GENERAL INFORMATION:
 ; APPLICANT: Meyerowitz, Elliott M.
 ; APPLICANT: Chang, Caren
 ; APPLICANT: Bleeker, Anthony B.
 ; TITLE OF INVENTION: PLANTS HAVING MODIFIED RESPONSE TO ETHYLENE
 ; NUMBER OF SEQUENCES: 34
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Richard F. Trecartin
 ; STREET: 3400 Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94111
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/530,010
 ; FILING DATE:
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/086,555
 ; FILING DATE: 01-JUL-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Trecartin, Richard F.
 ; REGISTRATION NUMBER: 31,801
 ; REFERENCE/DOCKET NUMBER: A-57515/RFT
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 781-1989
 ; TELEFAX: (415) 398-3249
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 738 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-530-010-7

Query Match 61.5%; Score 32; DB 1; Length 738;
 Best Local Similarity 62.5%; Pred. No. 1.6e+02;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 3 VVPXGXY 10
 ||| | |
 Db 498 VVPTGSFH 505

RESULT 12
 US-08-530-010-9
 ; Sequence 9, Application US/08530010
 ; Patent No. 5689055
 ; GENERAL INFORMATION:
 ; APPLICANT: Meyerowitz, Elliott M.
 ; APPLICANT: Chang, Caren
 ; APPLICANT: Bleeker, Anthony B.
 ; TITLE OF INVENTION: PLANTS HAVING MODIFIED RESPONSE TO ETHYLENE
 ; NUMBER OF SEQUENCES: 34
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Richard F. Trecartin
 ; STREET: 3400 Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94111
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/530,010
 ; FILING DATE:
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/086,555
 ; FILING DATE: 01-JUL-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Trecartin, Richard F.
 ; REGISTRATION NUMBER: 31,801
 ; REFERENCE/DOCKET NUMBER: A-57515/RFT
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 781-1989
 ; TELEFAX: (415) 398-3249
 ; INFORMATION FOR SEQ ID NO: 9:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 738 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-530-010-9

Query Match 61.5%; Score 32; DB 1; Length 738;
 Best Local Similarity 62.5%; Pred. No. 1.6e+02;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXY 10
 ||| | |
 Db 498 VVPTGSFH 505

RESULT 13
 US-08-530-010-11
 ; Sequence 11, Application US/08530010
 ; Patent No. 5689055
 ; GENERAL INFORMATION:
 ; APPLICANT: Meyerowitz, Elliott M.
 ; APPLICANT: Chang, Caren
 ; APPLICANT: Bleeker, Anthony B.
 ; TITLE OF INVENTION: PLANTS HAVING MODIFIED RESPONSE TO ETHYLENE
 ; NUMBER OF SEQUENCES: 34
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Richard F. Trecartin
 ; STREET: 3400 Embarcadero Center, Suite 3400

ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-57515-2/RFT
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 738 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-101B-3

Query Match 61.5%; Score 32; DB 1; Length 738;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
||| |
Db 498 VVPTGSHF 505

RESULT 14
US-08-484-101B-3
Sequence 3, Application US/08484101B
Patent No. 5824868
GENERAL INFORMATION:
APPLICANT: California Institute of Technology
TITLE OF INVENTION: PLANTS HAVING MODIFIED RESPONSE TO
ETHYLENE
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Richard F. Trecartin
STREET: 3400 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/08/484,101B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/
FILING DATE: 01-JUL-1994
CLASSIFICATION: 800
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 738 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-530-010-11

Query Match 61.5%; Score 32; DB 1; Length 738;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
||| |
Db 498 VVPTGSHF 505

RESULT 14
US-08-484-101B-3
Sequence 3, Application US/08484101B
Patent No. 5824868
GENERAL INFORMATION:
APPLICANT: California Institute of Technology
TITLE OF INVENTION: PLANTS HAVING MODIFIED RESPONSE TO
ETHYLENE
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Richard F. Trecartin
STREET: 3400 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/08/484,101B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/
FILING DATE: 01-JUL-1994
CLASSIFICATION: 800
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 738 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-101B-5

Query Match

61.5%; Score 32; DB 2; Length 738;

ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-57515-2/RFT
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 738 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-101B-3

Query Match 61.5%; Score 32; DB 2; Length 738;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
||| |
Db 498 VVPTGSHF 505

RESULT 15
US-08-484-101B-5
Sequence 5, Application US/08484101B
Patent No. 5824868
GENERAL INFORMATION:
APPLICANT: California Institute of Technology
TITLE OF INVENTION: PLANTS HAVING MODIFIED RESPONSE TO
ETHYLENE
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Richard F. Trecartin
STREET: 3400 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/08/484,101B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/
FILING DATE: 01-JUL-1994
CLASSIFICATION: 800
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 738 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-101B-5

Query Match

61.5%; Score 32; DB 2; Length 738;

Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
 | | | | |
Db 498 VVPTGSHF 505

Search completed: June 13, 2003, 14:14:31
Job time : 27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2003, 14:12:37 ; Search time 47 Seconds
(without alignments)
25.023 Million cell updates/sec

Title: US-09-909-164-42
Perfect score: 52
Sequence: 1 EEVFXGXHY 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	65.4	715	9	US-09-252-088-16
2	34	65.4	793	9	US-09-252-088-15
3	33	63.5	447	10	US-09-765-272-182
4	33	63.5	484	9	US-09-769-787-38
5	33	63.5	763	10	US-09-765-272-66
6	33	63.5	796	10	US-09-765-272-56
7	33	63.5	826	9	US-09-769-787-194
8	33	63.5	838	9	US-09-884-465A-8
9	33	63.5	840	9	US-09-884-465A-7
10	33	63.5	840	9	US-09-884-465A-10
11	33	63.5	1039	9	US-09-884-465A-6
12	32	61.5	1463	9	US-10-176-847-22
13	31	59.6	44	10	US-09-881-752A-292
14	31	59.6	387	10	US-09-815-242-10416
15	31	59.6	622	9	US-09-738-626-4919
16	31	59.6	653	9	US-09-820-843A-26
17	31	59.6	665	10	US-09-942-447-2
18	31	59.6	700	9	US-10-319-762-2
19	31	59.6	3472	9	US-10-027-806-4

20	31	59.6	3472	9	US-10-034-623-4
21	31	59.6	3472	9	US-10-027-801-4
22	30	57.7	7	9	US-09-909-062-1
23	30	57.7	7	9	US-09-909-062-9
24	30	57.7	24	9	US-09-909-062-130
25	30	57.7	72	10	US-09-925-297-664
26	30	57.7	121	9	US-09-852-797-68
27	30	57.7	121	9	US-09-852-797-85
28	30	57.7	121	10	US-09-853-161-68
29	30	57.7	121	10	US-09-853-161-85
30	30	57.7	121	10	US-09-852-659A-68
31	30	57.7	121	10	US-09-852-659A-85
32	30	57.7	135	9	US-09-982-598-359
33	30	57.7	135	9	US-09-989-293A-359
34	30	57.7	135	9	US-09-989-735-359
35	30	57.7	135	9	US-09-990-444-359
36	30	57.7	135	9	US-09-989-730-359
37	30	57.7	135	9	US-09-990-436-359
38	30	57.7	135	9	US-09-991-181-359
39	30	57.7	135	9	US-09-993-687-359
40	30	57.7	135	9	US-09-989-734-359
41	30	57.7	135	9	US-09-997-653-359
42	30	57.7	135	9	US-10-174-590-444
43	30	57.7	135	9	US-10-176-758-444
44	30	57.7	135	9	US-10-175-737-444
45	30	57.7	135	9	US-09-993-667-359

ALIGNMENTS

RESULT 1

US-09-252-088-16
; Sequence 16, Application US/09252088
; Publication No. US20030031682A1
; GENERAL INFORMATION:
; APPLICANT: BRODEUR, Bernard R.
; APPLICANT: RIOUX, Clment
; APPLICANT: BOYER, Martine
; APPLICANT: CHARLEBOIS, Isabelle
; APPLICANT: HAMEL, Jose
; APPLICANT: MARTIN, Denis
; TITLE OF INVENTION: NOVEL GROUP B STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 8331-9002
; CURRENT APPLICATION NUMBER: US/09/252,088
; CURRENT FILING DATE: 1999-02-18
; EARLIER APPLICATION NUMBER: US/60/075,425
; EARLIER FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 715
; TYPE: PRT
; ORGANISM: group B streptococcus
US-09-252-088-16

Query Match 65.4%; Score 34; DB 9; Length 715;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VWPXGXHY 10

Db 243 VVPHGDHY 250

RESULT 2

US-09-252-088-15
; Sequence 15, Application US/09252088
; Publication No. US20030031682A1
; GENERAL INFORMATION:
; APPLICANT: BRODEUR, Bernard R.
; APPLICANT: RIOUX, Clment
; APPLICANT: BOYER, Martine

; APPLICANT: CHARLEBOIS, Isabelle
 ; APPLICANT: HAMEL, Jose
 ; APPLICANT: MARTIN, Denis
 ; TITLE OF INVENTION: NOVEL GROUP B STREPTOCOCCUS ANTIGENS
 ; FILE REFERENCE: 8331-9002
 ; CURRENT APPLICATION NUMBER: US/09/252,088
 ; CURRENT FILING DATE: 1999-02-18
 ; EARLIER APPLICATION NUMBER: US/60/075,425
 ; EARLIER FILING DATE: 1998-02-20
 ; NUMBER OF SEQ ID NOS: 44
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 15
 ; LENGTH: 793
 ; TYPE: PRT
 ; ORGANISM: group B streptococcus
 US-09-252-088-15

Query Match 65.4%; Score 34; DB 9; Length 793;
 Best Local Similarity 75.0%; Pred. No. 1.6e+02;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
 DB 321 VVPHGDHY 328

RESULT 3
 US-09-765-272-182
 ; Sequence 182, Application US/09765272
 ; Patent No. US20020061545A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 ; NUMBER OF SEQUENCES: 452
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/765,272
 ; FILING DATE: 22-Jan-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/961,083
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brookes, A. Anders
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB340P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512
 ; INFORMATION FOR SEQ ID NO: 182:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 447 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 182:
 US-09-765-272-182

Query Match 63.5%; Score 33; DB 10; Length 447;
 Best Local Similarity 62.5%; Pred. No. 1.4e+02;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
 DB 178 IVPHGGHY 185

RESULT 4
 US-09-769-787-38
 ; Sequence 38, Application US/09769787
 ; Publication No. US20030091577A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Microbial Technics Limited
 ; APPLICANT: Gilbert, Christophe FG
 ; APPLICANT: Hansbro, Philip M
 ; TITLE OF INVENTION: Proteins
 ; FILE REFERENCE: PWC/P21129WO
 ; CURRENT APPLICATION NUMBER: US/09/769,787
 ; CURRENT FILING DATE: 2001-01-26
 ; PRIOR APPLICATION NUMBER: GB 9816337.1
 ; PRIOR FILING DATE: 1998-03-27
 ; PRIOR APPLICATION NUMBER: US 60/125164
 ; PRIOR FILING DATE: 1999-03-19
 ; NUMBER OF SEQ ID NOS: 388
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 38
 ; LENGTH: 484
 ; TYPE: PRT
 ; ORGANISM: Streptococcus pneumoniae
 US-09-769-787-38

Query Match 63.5%; Score 33; DB 9; Length 484;
 Best Local Similarity 62.5%; Pred. No. 1.5e+02;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
 DB 202 IVPHGGHY 209

RESULT 5
 US-09-765-272-66
 ; Sequence 66, Application US/09765272
 ; Patent No. US20020061545A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 ; NUMBER OF SEQUENCES: 452
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/765,272
 ; FILING DATE: 22-Jan-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/961,083
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brookes, A. Anders
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB340P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512

```

; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-09-765-272-66

Query Match      63.5%; Score 33; DB 10; Length 763;
Best Local Similarity 62.5%; Pred. No. 2.4e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      3 VVPXGXHY 10
Db      184 IVPXGDHY 191

RESULT 6
US-09-765-272-56
; Sequence 56, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 796 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 56:
US-09-765-272-56

Query Match      63.5%; Score 33; DB 10; Length 796;
Best Local Similarity 62.5%; Pred. No. 2.5e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      3 VVPXGXHY 10
Db      185 IVPXGDHY 192

RESULT 7

```

```

US-09-769-787-194
; Sequence 194, Application US/09769787
; Publication No. US20030091577A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129W0
; CURRENT APPLICATION NUMBER: US/09/769,787
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 194
; LENGTH: 826
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-787-194

Query Match      63.5%; Score 33; DB 9; Length 826;
Best Local Similarity 62.5%; Pred. No. 2.6e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      3 VVPXGXHY 10
Db      204 IVPXGDHY 211

RESULT 8
US-09-884-465A-8
; Sequence 8, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-8

Query Match      63.5%; Score 33; DB 9; Length 838;
Best Local Similarity 62.5%; Pred. No. 2.7e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      3 VVPXGXHY 10
Db      203 IVPXGDHY 210

RESULT 9
US-09-884-465A-7
; Sequence 7, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee

```

; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 840
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-7

Query Match 63.5%; Score 33; DB 9; Length 840;
Best Local Similarity 62.5%; Pred. No. 2.7e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
: || | ||
Db 205 IVPHGDIY 212

RESULT 10
US-09-884-465A-10
; Sequence 10, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 840
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-10

Query Match 63.5%; Score 33; DB 9; Length 840;
Best Local Similarity 62.5%; Pred. No. 2.7e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
: || | ||
Db 182 IVPHGDIY 189

RESULT 11
US-09-884-465A-6
; Sequence 6, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens

; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 1039
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-6

Query Match 63.5%; Score 33; DB 9; Length 1039;
Best Local Similarity 62.5%; Pred. No. 3.4e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
: || | ||
Db 202 IVPHGDIY 209

RESULT 12
US-10-176-847-22
; Sequence 22, Application US/10176847
; Publication No. US20030068636A1
; GENERAL INFORMATION:
; APPLICANT: Veiby, Petter Ole
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
; TITLE OF INVENTION: AND OVARIAN CANCER
; FILE REFERENCE: MRI-039
; CURRENT APPLICATION NUMBER: US/10/176,847
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 1463
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-176-847-22

Query Match 61.5%; Score 32; DB 9; Length 1463;
Best Local Similarity 45.5%; Pred. No. 7.6e+02;
Matches 5; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 BEVVPXGXHY 11
: || | ||
Db 438 EEVITSSHYT 448

RESULT 13
US-09-881-752A-292
; Sequence 292, Application US/09881752A
; Patent No. US20020115078A1
; GENERAL INFORMATION:
; APPLICANT: Kleenhouse, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020115078A1 Helicobacter Polypeptides in t
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/041002
; CURRENT APPLICATION NUMBER: US/09/881,752A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/833,457
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 292
; LENGTH: 44

; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-881-752A-292

Query Match 59.6%; Score 31; DB 10; Length 44;
Best Local Similarity 66.7%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 VVPXGXHS 11
||| |||
Db 34 VVPTGTHAS 42

RESULT 14

US-09-815-242-10416
; Sequence 10416, Application US/09815242
; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011a

; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 10416

; LENGTH: 387

; TYPE: PRT

; ORGANISM: Escherichia coli

US-09-815-242-10416

Query Match 59.6%; Score 31; DB 10; Length 387;

Best Local Similarity 50.08; Pred. No. 2.9e+02;

Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EVVPXGXH 9

||| |||

Db 188 EIIPTGGH 195

RESULT 15

US-09-738-626-4919

; Sequence 4919, Application US/09738626

; Publication No. US20020197605A1

GENERAL INFORMATION:

; APPLICANT: NAKAGAWA, SATOSHI

; APPLICANT: MIZOGUCHI, HIROSHI

; APPLICANT: ANDO, SEIKO

; APPLICANT: HAYASHI, MIKIRO

; APPLICANT: OCHIAI, KEIKO

; APPLICANT: YOKOI, HARUHIKO

; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4919
; LENGTH: 622
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4919

Query Match 59.6%; Score 31; DB 9; Length 622;

Best Local Similarity 50.0%; Pred. No. 4.8e+02;

Matches 6; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

Qy 1 EVVPXG--XHY 10

||:| | ||

Db 46 EEIPIGVFNHY 57

Search completed: June 13, 2003, 14:22:32

Job time : 48 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2003, 14:10:03 ; Search time 40 Seconds
(without alignments)
26.437 Million cell updates/sec

Title: US-09-909-164-42
Perfect score: 52
Sequence: 1 EEVVPKXGHYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	67.3	233	T02590	DNA binding protei
2	35	67.3	308	A72207	ftsH proteinase ac
3	35	67.3	360	E69086	cell division prot
4	35	67.3	1028	AF3286	ATP-dependent DNA
5	35	67.3	1396	S36851	L-shaped tail fibe
6	34	65.4	102	A42452	V1 protein - tobac
7	34	65.4	126	A25905	tachykinin B precu
8	34	65.4	279	C75538	hypothetical prote
9	34	65.4	495	T28717	hypothetical prote
10	34	65.4	534	A69284	coenzyme F420-quin
11	34	65.4	822	T46758	hypothetical 92.4K
12	34	65.4	1057	F89892	carbamoyl-phosphat
13	33	63.5	46	E97985	hypothetical prote
14	33	63.5	124	VKLJ51	trans-regulatory s
15	33	63.5	156	D82618	conserved hypothet
16	33	63.5	252	A2001	hypothetical prote
17	33	63.5	412	A48702	2-methyl-branched-
18	33	63.5	460	S69046	hypothetical prote
19	33	63.5	510	G86430	T518.1 protein - A
20	33	63.5	743	S38143	hypothetical prote
21	33	63.5	802	C95136	conserved domain p
22	33	63.5	819	B95136	hypothetical prote
23	33	63.5	828	E98004	conserved domain p
24	33	63.5	839	G95115	hypothetical prote
25	33	63.5	853	C97985	conserved hypothet
26	33	63.5	855	D98004	hypothetical prote
27	33	63.5	1039	H95115	histidine Motif-co
28	33	63.5	1039	D97985	hypothetical prote
29	32	61.5	162	A70939	probable monooxygen

30	32	61.5	225	2	S57810	hypothetical prote
31	32	61.5	286	2	C86169	prohibitin 2 [impo
32	32	61.5	311	2	H69194	GMP synthetase, su
33	32	61.5	322	2	AF3211	hypothetical prote
34	32	61.5	364	2	S47540	fructose-bisphosph
35	32	61.5	369	2	T17267	hypothetical prote
36	32	61.5	374	2	G59119	probable aspartate
37	32	61.5	425	2	T24111	hypothetical prote
38	32	61.5	454	2	G96568	probable non-photo
39	32	61.5	519	2	G84598	probable bZIP tran
40	32	61.5	536	2	AI3544	aldehyde dehydroge
41	32	61.5	627	2	A69663	DNA mismatch repai
42	32	61.5	696	2	A91247	phage transposase
43	32	61.5	701	2	S61239	hypothetical prote
44	32	61.5	716	1	JC5061	macrophage-stimula
45	32	61.5	738	2	A48246	ethylene-response

ALIGNMENTS

RESULT 1

T02590
DNA binding protein EREBP-2 - common tobacco
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 21-Jul-2000
C:Accession: T02590
R:Ohme-Takagi, M.; Shinshi, H.
Plant Cell 7, 173-182, 1995
A:Title: Ethylene-inducible DNA binding proteins that interact with an ethylene respo
A:Reference number: 214671; MUID:95276459; PMID:7756828
A:Accession: T02590
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-233 <OHN>
A:Cross-references: EMBL:D38126; NID:g790362; PIDN:BA07324.1; PID:g1208498
A:Experimental source: strain BY4; tissue-type leaf

Query Match 67.3%; Score 35; DB 2; Length 233;
Best Local Similarity 60.0%; Pred. No. 9.7;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 EEVVPKXGHY 10
: ||| | | |
Db 90 QAVVPKGRHY 99

RESULT 2

A72207
ftsH proteinase activity modulator HflK - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: A72207
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: A72207
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-308 <ARN>
A:Cross-references: GB:AE001819; GB:AE000512; NID:g4982396; PIDN:AA036885.1; PID:g498
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM1822
C:Superfamily: erythrocyte band 7 integral membrane protein

Query Match 67.3%; Score 35; DB 2; Length 308;
Best Local Similarity 75.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY      3 VVPXGXHY 10
      III I II
Db      41 VVPSGIHY 48

RESULT 3
E69086
cell division protein - Methanobacterium thermoautotrophicum (strain Delta H)
C:Species: Methanobacterium thermoautotrophicum
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999
C:Accession: E69086
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.;
K., S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A:Reference number: A69000; MUID:98037514; PMID:19371463
A:Accession: E69086
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-360 <MTH>
A:Cross-references: GB:AE000923; GB:AE000666; NID:g2622766; PIDN:AA886115.1; PID:g262277
A:Experimental source: strain Delta H
C:Genetics:
C:Gene: MTH1642
C:Superfamily: cell division protein MJ0174

      Query Match      67.3%; Score 35; DB 2; Length 360;
      Best Local Similarity 45.5%; Pred. No. 16;
      Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY      1 EEVVPXGXHY 11
      I::II I I::
Db      98 EDLVPMSGSH 108

RESULT 4
AF3286
ATP-dependent DNA helicase BMEI0275 [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AF3286
R:DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letes
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AB3252; PMID:11756688
A:Accession: AF3286
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1028 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL51457.1; PID:g17982167; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI0275
A:Map position: 1

      Query Match      67.3%; Score 35; DB 2; Length 1028;
      Best Local Similarity 54.5%; Pred. No. 48;
      Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      1 EEVVPXGXHY 11
      I::II I I I
Db      76 EKIVPPGARYS 86

RESULT 5
S36851
L-shaped tail fiber protein - phage T5
N:Alternate names: ltf protein
C:Species: phage T5
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 17-Nov-2000
C:Accession: S65934; S01984; S36851

      Query Match      65.4%; Score 34; DB 2; Length 102;
      Best Local Similarity 60.0%; Pred. No. 6.5;
      Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      2 EVVPXGXHY 11
      :::I I I I
Db      7 QVVPXGINYS 16

RESULT 7
A25905
tachykinin B precursor - bovine
N:Alternate names: neuromedin K
C:Species: Bos primigenius taurus (cattle)
C>Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 16-Jul-1999
C:Accession: A25905
R:Kotani, H.; Hoshimaru, M.; Nawa, H.; Nakanishi, S.
Proc. Natl. Acad. Sci. U.S.A. 83, 7074-7078, 1986
A:Title: Structure and gene organization of bovine neuromedin K precursor.
A:Reference number: A25905; MUID:86313713; PMID:3462746
A:Accession: A25905
A:Molecule type: DNA
A:Residues: 1-126 <KOT>
A:Cross-references: GB:M14351; NID:gl63587; PIDN:AAA30723.1; PID:gl63590
C:Superfamily: neurokinin B precursor
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-126/Product: protachykinin B #status predicted <MAT>

```

```

R:Kallman, A.V.; Kulshin, V.E.; Shlyapnikov, M.G.; Ksenzenko, V.N.; Kryukov, V.M.
FEBS Lett. 366, 46-48, 1995
A:Title: The nucleotide sequence of the bacteriophage T5 ltf gene.
A:Reference number: S65934; MUID:95309401; PMID:7789514
A:Accession: S65934
A:Molecule type: DNA
A:Residues: 1-1396 <KAL>
A:Cross-references: EMBL:X69460; NID:gi15415; PIDN:CAA49220.1; PID:gl15416
R:Kallman, A.V.; Kryukov, V.M.; Bayev, A.A.
Nucleic Acids Res. 16, 6230, 1988
A:Title: The nucleotide sequence of bacteriophage T5 DNA at the region between early
A:Reference number: S01982; MUID:88289370; PMID:3267228
A:Accession: S01984
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 934-985, 'A', 987-1396 <KA2>
A:Cross-references: EMBL:X07559
C:Genetics:
A:Gene: ltf
C:Keywords: late protein; tail fiber

      Query Match      67.3%; Score 35; DB 2; Length 1396;
      Best Local Similarity 50.0%; Pred. No. 67;
      Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      2 EVVPXGXHY 11
      :::I I I I
Db      1360 KTIAPGDHYS 1369

RESULT 6
A42452
V1 protein - tobacco yellow dwarf virus (strain Australia)
C:Species: tobacco yellow dwarf virus
C>Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
C:Accession: A42452
R:Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E.
Virology 187, 633-642, 1992
A:Title: The nucleotide sequence of the infectious cloned DNA component of tobacco ye
A:Reference number: A42452; MUID:92188538; PMID:1546458
A:Accession: A42452
A:Molecule type: DNA
A:Residues: 1-102 <MOR>
A:Cross-references: GB:M81103; NID:g335283; PIDN:AAA47947.1; PID:g335284

      Query Match      65.4%; Score 34; DB 2; Length 102;
      Best Local Similarity 60.0%; Pred. No. 6.5;
      Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      2 EVVPXGXHY 11
      :::II I I I
Db      7 QVVPXGINYS 16

RESULT 7
A25905
tachykinin B precursor - bovine
N:Alternate names: neuromedin K
C:Species: Bos primigenius taurus (cattle)
C>Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 16-Jul-1999
C:Accession: A25905
R:Kotani, H.; Hoshimaru, M.; Nawa, H.; Nakanishi, S.
Proc. Natl. Acad. Sci. U.S.A. 83, 7074-7078, 1986
A:Title: Structure and gene organization of bovine neuromedin K precursor.
A:Reference number: A25905; MUID:86313713; PMID:3462746
A:Accession: A25905
A:Molecule type: DNA
A:Residues: 1-126 <KOT>
A:Cross-references: GB:M14351; NID:gl63587; PIDN:AAA30723.1; PID:gl63590
C:Superfamily: neurokinin B precursor
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-126/Product: protachykinin B #status predicted <MAT>

```


Query Match 65.4%; Score 34; DB 2; Length 126;
 Best Local Similarity 66.7%; Pred. No. 8.2;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGXH 9
 I:| | | |
 Db 28 EQVPGGGH 36

RESULT 8

C7538
 hypothetical protein - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000
 C:Accession: C7538
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 S.; Smith, H.O.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896; PMID:10567266
 A:Accession: C7538
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-279 <WHI>
 A:Cross-references: GB:AE001889; GB:AE000513; NID:g6457944; PIDN:AAF09867.1; PID:g645795
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR0271
 A:Map position: 1
 C:Superfamily: Deinococcus radiodurans hypothetical protein DR0271

Query Match 65.4%; Score 34; DB 2; Length 279;
 Best Local Similarity 75.0%; Pred. No. 19;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 VPXGXHYS 11
 I: | | | |
 Db 100 VPLGRHYS 107

RESULT 9

T28717
 hypothetical protein F10D2.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T28717
 R:Graves, T.; Wohldmann, P.; Gillam, B.
 submitted to the EMBL Data Library, September 1997
 A:Description: The sequence of C. elegans cosmid F10D2.
 A:Reference number: Z20515
 A:Accession: T28717
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-495 <GRA>
 A:Cross-references: EMBL:AF022972; PIDN:AAC48234.1; GSPDB:GN000023; CBSP:F10D2.3
 A:Experimental source: strain Bristol N2; clone F10D2
 C:Genetics:
 A:Gene: CBSP:F10D2.3
 A:Map position: 5
 A:Introns: 37/2; 90/1; 113/1; 183/3; 356/3; 380/3; 428/2

Query Match 65.4%; Score 34; DB 2; Length 495;
 Best Local Similarity 50.0%; Pred. No. 36;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGXHY 10
 I: | | | |
 Db 218 ENIVPTGKH 227

RESULT 10

A69284

coenzyme F420-quinone oxidoreductase (EC 1.6.5.-) 56K chain - Archaeoglobus fulgidus
 N:Alternate names: sarcosine oxidase alpha chain soxa homolog
 C:Species: Archaeoglobus fulgidus
 C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 19-May-2000
 C:Accession: A69284; S45665
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
 : Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
 Nature 390, 364-370, 1997
 A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes,
 Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
 A:Reference number: A69250; MUID:98049343; PMID:9389475

A:Accession: A69284
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-534 <KLE>
 A:Cross-references: GB:AF001086; GB:AF000782; NID:g2689409; PIDN:AAB90959.1; PID:g265
 R:Kunow, J.; Linder, D.; Stetter, K.O.; Thauer, R.K.
 Eur. J. Biochem. 223, 503-511, 1994

A:Title: F(420)H(2): quinone oxidoreductase from Archaeoglobus fulgidus. Characteriza
 A:Reference number: S45665; MUID:94333340; PMID:8055920
 A:Accession: S45665
 A:Molecule type: protein
 A:Residues: 2,'X','A','X','6-7,'X','9-10,'XX',13-14 <KUN>
 A:Note: the authors could not distinguish between glutamate and cysteine for residues
 C:Superfamily: glutamate synthase small chain
 C:Keywords: oxidoreductase

Query Match 65.4%; Score 34; DB 2; Length 534;
 Best Local Similarity 45.5%; Pred. No. 39;
 Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYS 11
 : : | | | |
 Db 119 DKVFPFGSHYT 129

RESULT 11

T46758
 hypothetical 92.4K protein - Streptococcus agalactiae
 C:Species: Streptococcus agalactiae
 C>Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 21-Jul-2000
 C:Accession: T46758
 R:Spellerberg, B.; Rozdzinski, E.; Martin, S.; Weber-Heymann, J.; Schnitzler, N.; L
 Infect. Immun. 67, 871-878, 1999
 A:Title: Lmb, a protein with similarities to the Lrai adhesin family, mediates attach
 A:Reference number: Z24091; MUID:99115568; PMID:9916102
 A:Accession: T46758
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-822 <SPE>
 A:Cross-references: EMBL:AF062533; NID:g4249622; PIDN:AAD13797.1; PID:g4249624
 A:Experimental source: strain R268
 C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 65.4%; Score 34; DB 2; Length 822;
 Best Local Similarity 75.0%; Pred. No. 62;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
 I: | | | |
 Db 350 VVPXGDHY 357

RESULT 12

F89892
 carbamoyl-phosphate synthase large chain [imported] - Staphylococcus aureus (strain N
 C:Species: Staphylococcus aureus
 C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
 C:Accession: F89892
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
 ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K

C.: Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.

A:Reference number: A89758; MUID:21311952; PMID:11418146

A:Accession: F89892

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1057 <KUR>

A:Cross-references: GB:BA000018; PID:g13701002; PIDN:BA042298.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: pyrAB

C:Superfamily: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain; biotin

Query Match 65.4%; Score 34; DB 2; Length 1057;

Best Local Similarity 60.0%; Pred. No. 81;

Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVPXGXHY 11

I:| | | |

Db 190 EIVSNGLHY 199

RESULT 13

E97985

hypothetical protein spr0909 [imported] - *Streptococcus pneumoniae* (strain R6)

C:Species: *Streptococcus pneumoniae*

C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001

C:Accession: E97985

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E

y, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M

y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.

A:Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.

A:Reference number: A97872; MUID:21428245; PMID:11544234

A:Accession: E97985

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-46 <KUR>

A:Cross-references: GB:AE007317; PIDN:AAK99713.1; PID:g15458516; GSPDB:GN00174

C:Genetics:

A:Gene: spr0909

Query Match 63.5%; Score 33; DB 2; Length 46;

Best Local Similarity 62.5%; Pred. No. 4.5;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10

:| | | |

Db 21 IVPHGGHY 28

RESULT 14

VKLJST

trans-regulatory splicing protein - simian immunodeficiency virus SIVcpz

N:Alternate names: anti-repression trans-activator; art protein; rev protein; trs prote

C:Species: simian immunodeficiency virus SIVcpz

A:Note: host Pan troglodytes (chimpanzee)

C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jul-1999

C:Accession: S09988

R:Huet, T.; Cheynier, R.; Meyerhans, A.; Roelants, G.; Wain-Hobson, S.

Nature 345, 356-359, 1990

A:Title: Genetic organization of a chimpanzee lentivirus related to HIV-1.

A:Reference number: S09983; MUID:90259077; PMID:2188136

A:Accession: S09988

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-124 <HUE>

A:Cross-references: EMBL:X52154; NID:g58866; PIDN:CAA36405.1; PID:g763085

C:Genetics:

A:Gene: rev; trs; art

A:Introns: 27/1

C:Superfamily: AIDS trans-regulatory splicing protein

C:Keywords: AIDS; immunodeficiency; splicing protein; transcription regulation

Query Match 63.5%; Score 33; DB 1; Length 124;

Best Local Similarity 60.0%; Pred. No. 13;

Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVPXGXHY 11

I:| | | |

Db 107 EIVPAGGNY 116

RESULT 15

D82618

conserved hypothetical protein XF1950 [imported] - *Xylella fastidiosa* (strain 9a5c)

C:Species: *Xylella fastidiosa*

C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C:Accession: D82618

R:anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Seq

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.

A:Reference number: A82515; MUID:20365717; PMID:10910347

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: D82618

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-156 <SIM>

A:Cross-references: GB:AE004014; GB:AE003849; NID:g9107044; PIDN:AAF84752.1; GSPDB:GN

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carr

as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima J.P.; Krieger, J.E.; Kuramae, E.E.; La

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins

, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Miyaki, C.

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv

M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF1950

Query Match 63.5%; Score 33; DB 2; Length 156;

Best Local Similarity 55.6%; Pred. No. 17;

Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGXH 9

I:| | | |

Db 119 EEILPQGVH 127

Search completed: June 13, 2003, 14:12:33

Job time : 41 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2003, 14:10:03 ; Search time 22 Seconds
(without alignments)
20.738 Million cell updates/sec

Title: US-09-909-164-42
Perfect score: 52
Sequence: 1 EEWVFXGXHYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	35	67.3	1058	1	CARB_FUSNN
2	35	67.3	1396	1	VLFF_BPT5
3	34	65.4	102	1	YLIK_TYDVA
4	34	65.4	126	1	TKNK_BOVIN
5	34	65.4	1057	1	CARB_STAAM
6	34	65.4	1057	1	CARB_STAAM
7	33	63.5	124	1	REV_SIVCZ
8	33	63.5	460	1	UMEL_YEAST
9	33	63.5	743	1	YK47_YEAST
10	32	61.5	308	1	GAAB_METH
11	32	61.5	363	1	ALPB_SHEEP
12	32	61.5	627	1	MOTL_BACSU
13	32	61.5	735	1	ETRI_BRAOL
14	32	61.5	738	1	ETRI_ARATH
15	32	61.5	906	1	CENC_MOUSE
16	32	61.5	1374	1	YQ56_CAEEL
17	32	61.5	1378	1	RON_MOUSE
18	31	59.6	130	1	SZ05_RAT
19	31	59.6	212	1	MSRA_VIBCH
20	31	59.6	319	1	YHAL_CRYPA
21	31	59.6	363	1	ALPB_HUMAN
22	31	59.6	363	1	ALPB_RABIT
23	31	59.6	387	1	THIK_ECOLI
24	31	59.6	426	1	SLSI_YARLI
25	31	59.6	427	1	TOLB_HAEIN
26	31	59.6	991	1	DHPL_SCHPO
27	31	59.6	1188	1	KPBA_CAEEL
28	31	59.6	1410	1	PDBB_YEAST
29	31	59.6	1499	1	ALOC_HUMAN
30	31	59.6	3099	1	POLG_PEMVM
31	30.5	58.7	268	1	ILIA_BOVIN
32	30.5	58.7	268	1	ILIA_CAPHI
33	30.5	58.7	268	1	ILIA_SHEEP

34 30 57.7 121 1 TKNK_HUMAN
35 30 57.7 152 1 Y16D_BPT4
36 30 57.7 160 1 Y201_METJA
37 30 57.7 223 1 COAT_CTV36
38 30 57.7 232 1 SCOA_HELPJ
39 30 57.7 232 1 SCOA_HELPJ
40 30 57.7 316 1 COQ3_YEAST
41 30 57.7 328 1 RPOA_BORPE
42 30 57.7 353 1 T2BA_BACAR
43 30 57.7 386 1 TI0D_HUMAN
44 30 57.7 394 1 NUCC_SYNY3
45 30 57.7 427 1 TRB1_AERPE

ALIGNMENTS

RESULT 1

CARB_FUSNN STANDARD; PRT; 1058 AA.
AC Q8RG86;
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).
DE Carbamoyl-phosphate synthetase ammonia chain.
GN CARB OR FN0422.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A., Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L., Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A., Larsen N., D'Souza M., Walunas T., Busch G., Hasekorn R., Foust M., Kyrides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium nucleatum strain ATCC 25586.";
RT J. Bacteriol. 184:2005-2018(2002).
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.
CC -!- COFACTOR: Binds three manganese ions (By similarity).
CC -!- PATHWAY: Arginine biosynthesis.
CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity).
CC -!- SIMILARITY: BELONGS TO THE CARB FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; AE010554; AAL94625.1; ALT_INIT.
InterPro; IPR005483; CPase_L.
InterPro; IPR005479; CPase_L_D2.
InterPro; IPR005480; CPase_L_D3.
InterPro; IPR005481; CPase_L_N.
InterPro; IPR004362; MGS_like.
Pfam; PF00289; CPase_L_chain; 2.
Pfam; PF02786; CPase_L_D2; 2.
Pfam; PF02787; CPase_L_D3; 1.
Pfam; PF02142; MGS; 1.
PRINTS; PR00098; CPASE.
PROSITE; PS00866; CPASE_1; 2.

DR PROSITE; PS00867; CPSASE_2; 2.
 KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
 ATP-binding; Manganese; Complete proteome.
 FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN
 FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 930 1058 ALLOSTERIC DOMAIN.
 FT REPEAT 1 546
 FT REPEAT 547 1058
 FT NP_BIND 153 210
 FT NP_BIND 302 352
 FT NP_BIND 284 284
 FT METAL 298 298
 FT METAL 300 300
 FT METAL 820 820
 FT METAL 832 832
 FT METAL 832 832
 SQ SEQUENCE 1058 AA; 117451 MW; ED7037AF77C1E39F CRC64;
 Query Match 67.3%; Score 35; DB 1; Length 1058;
 Best Local Similarity 60.0%; Pred. No. 20;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 2 EVVPGXHYHS 11
 Db 190 EIVPGLNYS 199
 RESULT 2
 VLTF_BPT5 STANDARD; PRT; 1396 AA.
 AC P13390; O48502;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE L-shaped tail fiber protein (LTF protein).
 GN LTF.
 OS Bacteriophage T5.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
 OC T5-like viruses.
 OX NCBI_TaxID=10726;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95309401; PubMed=7789514;
 RA Kallman A.V., Kulshin V.E., Shlyapnikov M.G., Ksenzenko V.N.,
 RT Kryukov V.M.;
 RT "The nucleotide sequence of the bacteriophage T5 ltf gene.";
 RL FEBS Lett. 366:46-48(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Kallman A.V.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
 RX MEDLINE=88289370; PubMed=3267228;
 RA Kallman A.V., Kryukov V.M., Bayev A.A.;
 RT "The nucleotide sequence of bacteriophage T5 DNA at the region
 between early and late genes.";
 RL Nucleic Acids Res. 16:6230-6230(1988).
 CC -!- FUNCTION: NONESSENTIAL PROTEIN THAT MEDIATES BINDING TO THE
 CC POLYMANNANOSE O ANTIGEN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X69460; CAA49220.1; -
 CC EMBL; AJ001191; CAA04591.1; -
 DR PIR; S01982; S01982.
 KW Late protein.

FT CONFLICT 986 986 V -> A (IN REF. 2).
 SQ SEQUENCE 1396 AA; 147989 MW; 18CD2192F65FFFC1 CRC64;
 Query Match 67.3%; Score 35; DB 1; Length 1396;
 Best Local Similarity 50.0%; Pred. No. 26;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 2 EVVPGXHYHS 11
 Db 1360 KTIAPGDHYS 1369
 RESULT 3
 Y1LK_TYDVA STANDARD; PRT; 102 AA.
 AC P31619;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-OCT-1993 (Rel. 27, Last annotation update)
 DE Hypothetical 11.2 kDa protein.
 GN Y1.
 OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
 OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
 OX NCBI_TaxID=31599;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92188538; PubMed=1546458;
 RA Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.E.;
 RT "The nucleotide sequence of the infectious cloned DNA component of
 RT tobacco yellow dwarf virus reveals features of geminiviruses
 RT infecting monocotyledonous plants.";
 RL Virology 187:633-642(1992).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M81103; AAA47947.1; -
 DR PIR; A42452; A42452.
 DR InterPro; IPR002621; Gemini_mov.
 DR Pfam; PF01708; Gemini_mov; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 102 AA; 11178 MW; A40ECF1E0AF55B67 CRC64;
 Query Match 65.4%; Score 34; DB 1; Length 102;
 Best Local Similarity 60.0%; Pred. No. 2.7;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 2 EVVPGXHYHS 11
 Db 7 QVVPSCINYS 16
 RESULT 4
 TKNK_BOVIN STANDARD; PRT; 126 AA.
 ID P08836;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neurokinin B precursor (NKB) (Neuromedin K).
 GN TAC3 OR NKB.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.

```

RX MEDLINE=86313713; PubMed=3462746;
RA Kotani H., Hoshimaru M., Nawa H., Nakanishi S.;
RT "Structure and gene organization of bovine neurexin K precursor.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:7074-7078(1986).
CC -|- FUNCTION: TACHYKININS ARE ACTIVE PEPTIDES WHICH EXCITE NEURONS,
CC EVOKE BEHAVIORAL RESPONSES, ARE POTENT VASODILATORS AND
CC SECRETAGOGUES, AND CONTRACT (DIRECTLY OR INDIRECTLY) MANY SMOOTH
CC MUSCLES.
CC -|- SUBCELLULAR LOCATION: Secreted.
CC -|- SIMILARITY: BELONGS TO THE TACHYKININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M14351; AAA30723.1; JOINED.
DR EMBL; M14347; AAA30723.1; JOINED.
DR EMBL; M14348; AAA30723.1; JOINED.
DR EMBL; M14349; AAA30723.1; JOINED.
DR EMBL; M14350; AAA30723.1; JOINED.
DR PIR; A25905; A25905.
DR InterPro; IPR003635; Neurokinin.
DR InterPro; IPR002040; Tachykinin.
DR ProDom; PD020370; Neurokinin; 1.
DR PROSITE; PS00267; Tachykinin; 1.
KW Tachykinin; Neuropeptide; Cleavage on pair of basic residues;
KW Amidation; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT PROPEP 21 83
FT PEPTIDE 86 95
FT PROPEP 99 126
FT MOD_RES 95 95
FT SEQUENCE 126 AA; 13871 MW; 446EF43498EC059 CRC64;

Query Match 65.4%; Score 34; DB 1; Length 126;
Best Local Similarity 66.7%; Pred. No. 3.4;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGXH 9
DB 28 EQVPGGGH 36
I:|:|:|

RESULT 5
CARB_STAAM STANDARD; PRT; 1057 AA.
AC Q99UR5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
DE phosphate synthetase ammonia chain).
GN CARB OR PYRAB OR SAV1203 OR SA1046.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mu50 / ATCC 700699, and N315;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-O., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus

```

```

aureus.";
RL Lancet 357:1225-1240(2001).
CC -|- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O -> 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -|- COFACTOR: Binds three manganese ions (By similarity).
CC -|- PATHWAY: Arginine biosynthesis.
CC -|- PATHWAY: Pyrimidine biosynthesis; first step.
CC -|- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -|- SIMILARITY: BELONGS TO THE CARB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP003361; BAB57365.1; -.
DR EMBL; AP003132; BAB42298.1; -.
DR HSP; P00968; 1CS0.
DR InterPro; IPR005483; CPase_L.
DR InterPro; IPR005479; CPase_L_D2.
DR InterPro; IPR005480; CPase_L_D3.
DR InterPro; IPR005481; CPase_L_N.
DR InterPro; IPR004362; MGS-like.
DR Pfam; PF00289; CPase_L_chain; 2.
DR Pfam; PF02786; CPase_L_D2; 2.
DR Pfam; PF02787; CPase_L_D3; 1.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPASE.
DR PROSITE; PS00866; CPASE_1; 2.
DR PROSITE; PS00867; CPASE_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1057 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1057
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1057 AA; 117171 MW; E3E179EF0591F0F8 CRC64;

Query Match 65.4%; Score 34; DB 1; Length 1057;
Best Local Similarity 60.0%; Pred. No. 32;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVPGXGHS 11
DB 190 EIVSNGLHYS 199
I:|:|:|

RESULT 6
CARB_STAAM STANDARD; PRT; 1057 AA.
AC P58940;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
DE phosphate synthetase ammonia chain).
GN CARB OR PYRAB OR MW1086.
OS Staphylococcus aureus (strain MW2).

```

CC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=196620;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22040717; PubMed=12044378;
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
 RA Nagai Y., Iwama N., Asano K., Naito T., Kuroda H., Cui L.,
 RA Yamamoto K., Hiramatsu K.,
 RT "Genome and virulence determinants of high virulence community-
 acquired MRSA";
 RL Lancet 359:1819-1827(2002).
 CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O -> 2 ADP +
 phosphate + L-glutamate + carbamoyl phosphate.
 CC -!- COFACTOR: Binds three manganese ions (by similarity).
 CC -!- PATHWAY: Arginine biosynthesis.
 CC -!- PATHWAY: Pyrimidine biosynthesis.
 CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain
 promotes the hydrolysis of glutamine to ammonia, which is used by
 the large (or ammonia) chain to synthesize carbamoyl phosphate (by
 similarity).
 CC -!- SIMILARITY: BELONGS TO THE CARB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AP004825; BAB94951.1;
 DR PROSITE; PS00866; CPSASE_1; 2.
 DR PROSITE; PS00867; CPSASE_2; 2.
 KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
 KW ATP-binding; Manganese.
 FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
 FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 930 1057 ALLOSTERIC DOMAIN.
 FT REPEAT 1 546
 FT REPEAT 547 1057
 FT NP_BIND 153 210 ATP (POTENTIAL).
 FT NP_BIND 302 352 ATP (POTENTIAL).
 FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
 FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
 FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
 FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
 FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
 SQ SEQUENCE 1057 AA; 117185 MW; D8E3B09F9BC6F152 CRC64;
 Query Match 65.4%; Score 34; DB 1; Length 1057;
 Best Local Similarity 60.0%; Pred. No. 32;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 2 EVVPXGXHYS 11
 DB 190 EIVSNGLHYS 199
 RESULT 7
 REV_SIVCZ
 ID REV_SIVCZ STANDARD; PRT; 124 AA.
 AC P17280;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE REV protein (Anti-repression transactivator protein) (ART/TRS).
 GN REV.
 OS Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
 OC Viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11723;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE=90259077; PubMed=2188136;
 RT Huet T., Cheynier R., Meyerhans A., Roelants G., Wain-Hobson S.;
 RA "Genetic organization of a chimpanzee lentivirus related to HIV-1.";
 RL Nature 345:356-359(1990).
 CC -!- FUNCTION: REV APPEARS TO ACT POST-TRANSCRIPTIONALLY TO RELIEVE
 CC NEGATIVE REPRESSION OF GAG AND ENV PRODUCTION.
 CC -!- SUBCELLULAR LOCATION: Nuclear; accumulates in the nucleoli.
 CC -!- PTM: PHOSPHOPROTEIN WHOSE STATE OF PHOSPHORYLATION IS MEDIATED
 CC BY A SPECIFIC SERINE KINASE ACTIVITY PRESENT IN THE NUCLEUS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X52154; CAA36405.1;
 DR PIR; S09988; VKLJSI.
 DR HIV; X52154; REVSCPZ.
 DR InterPro; IPR000625; REV_protein.
 DR Pfam; PF00424; REV; 1.
 KW Transcription regulation; AIDS; Phosphorylation; Nuclear protein.
 SQ SEQUENCE 124 AA; 13701 MW; F5877D1BDF65A7B2 CRC64;
 Query Match 63.5%; Score 33; DB 1; Length 124;
 Best Local Similarity 60.0%; Pred. No. 5.5;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 2 EVVPXGXHYS 11
 DB 107 ETVPAGGNY 116
 RESULT 8
 UMEI_YEAST
 ID UMEI_YEAST STANDARD; PRT; 460 AA.
 AC Q03010; P87330;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Meiosis negative regulator UME1.
 GN UME1 OR WTH3 OR YPL139C OR LP17C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A364A;
 RA Mallory M.J., Strich R.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / AB972;
 RX MEDLINE=97313271; PubMed=9169875;
 RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansong W.,
 RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
 RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
 RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
 RA Dietrich F.S., Dilius H., Dipaolo T., Dubois E., Duesterhoeft A.,
 RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
 RA Hall J., Hebling U., Heumann K., Hilbert H., Hallier L.,
 RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Kleene K.,
 RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
 RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
 RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
 RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Schafie M.,
 RA Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,
 RA Urestrazu L.A., Ushinsky S., Vierdeels F., Viessers S., Voss H.,
 RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
 RA Zhong W.W., Zollner A., Vo D.H., Hani J.;

RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
RL Nature 387:103-105(1997).
CC -!- FUNCTION: TRANSCRIPTIONAL MODULATOR WITH ROLES IN MELIORIC
CC REGULATION AND SILENCING. NEGATIVE REGULATOR OF MEIOSIS.
CC -!- SIMILARITY: CONTAINS 4 WD REPEATS (TRP-ASP DOMAINS).
CC -!- SIMILARITY: STRONG, TO YEAST WTM1 AND WTM2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U10280; AAB40937.1; -;
DR EMBL; U43703; AAB68221.1; -;
DR TRANSFAC; T04309; -;
DR SGD; S0006060; UME1.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 3.
DR SMART; SM00320; WD40; 3.
DR PROSITE; PS00678; WD_REPEATS_1; FALSE_NEG.
DR PROSITE; PS00082; WD_REPEATS_2; FALSE_NEG.
DR PROSITE; PS0294; WD_REPEATS_REGION; FALSE_NEG.
KW Transcription regulation; Meiosis; Repeat; WD repeat.
FT REPEAT 233 271 WD 1.
FT REPEAT 276 316 WD 2.
FT REPEAT 339 379 WD 3.
FT REPEAT 411 451 WD 4.
FT REPEAT 411 451 WD 4.
SQ SEQUENCE 460 AA; 51022 MW; AA6F60448B7BCBA9 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 460;
Best Local Similarity 62.5%; Pred. No. 22;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
:|:| | |
DB 85 IVPLGLHY 92

RESULT 9
YK47_YEAST STANDARD; PRT; 743 AA.
AC P36148;
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE Hypothetical 83.6 kDa protein in CCPI-MET1 intergenic region.
GN YK067W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RC STRAIN=S288C;
RA van Vliet-Reedijk J.C., Planta R.J.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- SIMILARITY: STRONG, TO YEAST YBL011W.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z28292; CAA82146.1; -;
DR PIR; S38143; S38143.
DR SGD; S0001775; YK067W.

DR InterPro; IPR002123; Acyltransferase.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 31 55 POTENTIAL.
FT TRANSMEM 69 85 POTENTIAL.
FT TRANSMEM 502 524 POTENTIAL.
FT TRANSMEM 539 555 POTENTIAL.
SQ SEQUENCE 743 AA; 83644 MW; 84B9946E56B82F15 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 743;
Best Local Similarity 75.0%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
:|:| | |
DB 294 VVPCGLHY 301

RESULT 10
GAAB_METTH STANDARD; PRT; 308 AA.
AC Q26806;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE GMP synthase [glutamine-hydrolyzing] subunit B (EC 6.3.5.2) (GMP
DE synthetase).
GN GAAB OR MTH710.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriacea; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwan N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
CC -!- CATALYTIC ACTIVITY: ATP + xanthosine 5'-phosphate + L-glutamine +
CC H(2)O = AMP + diphosphate + GMP + L-glutamate.
CC -!- PATHWAY: GMP biosynthesis.
CC -!- SUBUNIT: HETERODIMER COMPOSED OF A GLUTAMINE AMIDOTRANSFERASE
CC SUBUNIT (A) AND A GMP SYNTHASE SUBUNIT (B) (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE GMP SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000850; AAB85215.1; ALT_INIT.
DR HSSP; P04079; IGPM.
DR InterPro; IPR001674; GMP_synt_C.
DR Pfam; PF00958; GMP_synt_C; 1.
DR TIGRfams; TIGR00884; quaa_Cterm; 1.
KW Ligase; GMP biosynthesis; Purine biosynthesis; ATP-binding;
KW Complete proteome.
FT DOMAIN 33 184 GMP-BINDING (BY SIMILARITY).
FT NP_BIND 29 35 ATP (BY SIMILARITY).
SQ SEQUENCE 308 AA; 34403 MW; F2DCF6ED202CABC1 CRC64;

Query Match 61.5%; Score 32; DB 1; Length 308;
Best Local Similarity 63.8%; Pred. No. 23;

Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 EEWVPGXGHS 11
||||| | | |
Db 216 EEWVSGHLHS 226

RESULT 11

ID ALFB SHEEP STANDARD; PRT; 363 AA.
AC P52210;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Fructose-bisphosphate aldolase B (EC 4.1.2.13) (Liver-type aldolase).
GN ALDOB.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Wesonephros;
RX MEDLINE=94368863; PubMed=8086469;
RA Gianquinto L., Pailhoux E.A., Bezaud J., Servel N., Kirszenbaum M.,
Cotinot C.;
RT "Cloning and characterization of a full-length cDNA coding for ovine
aldolase B from fetal mesonephros.";
RL Biochim. Biophys. Acta 1219:223-227(1994).
CC -1- CATALYTIC ACTIVITY: D-fructose 1,6-bisphosphate = glycerone
phosphate + D-glyceraldehyde 3-phosphate.
CC -1- PATHWAY: Glycolysis; sixth step.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- MISCELLANEOUS: IN VERTEBRATES, THREE FORMS OF THIS UBIQUITOUS
GLYCOLYTIC ENZYME ARE FOUND, ALDOLASE A IN MUSCLE, ALDOLASE B IN
LIVER, & ALDOLASE C IN BRAIN.
CC -1- SIMILARITY: BELONGS TO CLASS I FRUCTOSE-BISPHOSPHATE ALDOLASE
FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; 229372; CAA82563.1; -
HSP; P00883; IADO.
InterPro; IPR000741; Aldolase_I.
DR Pfam; PF00274; glycolytic_enz; 1.
DR ProDom; PD001128; Aldolase_I; 1.
DR PROSITE; PS00158; ALDOLASE_CLASS_I; 1.
KW Lyase; Schiff base; Glycolysis; Multigene family.
FT INIT_MET 0
FT BINDING 55 55 C-1-PHOSPHATE GROUP OF THE SUBSTRATE.
FT BINDING 146 146 C-1-PHOSPHATE GROUP OF THE SUBSTRATE.
FT BINDING 229 229 SCHIFF-BASE WITH DIHYDROXYACETONE-P.
FT ACT_SITE 363 363 ESSENTIAL FOR ENHANCED ACTIVITY OF THE
ENZYME TOWARD FRUCTOSE 1,6-BISPHOSPHATE
AS COMPARED WITH FRUCTOSE 1-PHOSPHATE.
SQ SEQUENCE 363 AA; 39500 MW; FC8B4566821E2BD CRC64;

Query Match 61.5%; Score 32; DB 1; Length 363;
Best Local Similarity 62.5%; Pred. No. 28;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EEWVPGXGHS 9
||||| | | |
Db 189 EEWVPGSH 196

RESULT 12

ID MUTL_BACSU STANDARD; PRT; 627 AA.
AC P49850;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA mismatch repair protein mutL.
GN MUTL.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96349107; PubMed=8760914;
RA Ginetti F., Perego M., Albertini A.M., Galizzi A.;
RT "Bacillus mutL muts mutL operon: identification, nucleotide
sequence and mutagenesis.";
RL Microbiology 142:2021-2029(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
Brouillet S., Bruchi C.V., Caldwell B., Capuano V., Carter N.M.,
Choi S.K., Codani J.J., Connerth I.F., Cummings N.J., Daniel R.A.,
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
Etian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
Joris B., Karamata D., Kasahara Y., Koningsstein G., Krogh S., Kumano M.,
Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
Kurita K., Lapidus A., Lardiniois S., Lauber J., Lazarevic V.,
Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadale Y.,
Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis.";
RL Nature 390:249-256(1997).
CC -1- FUNCTION: THIS PROTEIN IS INVOLVED IN THE REPAIR OF MISMATCHES IN
DNA. IT IS REQUIRED FOR DAM-DEPENDENT METHYL-DIRECTED DNA MISMATCH
REPAIR. MAY ACT AS A "MOLECULAR MATCHMAKER". A PROTEIN THAT
PROMOTES THE FORMATION OF A STABLE COMPLEX BETWEEN TWO OR MORE
DNA-BINDING PROTEINS IN AN ATP-DEPENDENT MANNER WITHOUT ITSELF
BEING PART OF A FINAL EFFECTOR COMPLEX (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTL/HEXB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U27343; AAB19236.1; -
EMBL; Z99112; CAB13578.1; -


```

DR HSP; P23367; IBKN.
DR Subtilist; BG11402; mutL.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR002099; DNA_mis_repair.
DR InterPro; IPR004359; His_Kin_sig.
DR Pfam; PF01119; DNA_mis_repair; 1.
DR Pfam; PF02518; HATPase_C; 1.
DR TIGRFAMs; TIGR00585; mutL; 1.
DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
KW DNA repair; Complete proteome.
SQ SEQUENCE 627 AA; 70431 MW; 068A0509CC255343 CRC64;

Query Match 61.5%; Score 32; DB 1; Length 627;
Best Local Similarity 54.5%; Pred. No. 49;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 BEVVPXGHYS 11
| : : : :
Db 488 EMIVPLTTHYS 498

RESULT 13
ETRL_BRAOL
ID ETRL_BRAOL STANDARD; PRT; 735 AA.
AC 049230;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Ethylene receptor (EC 2.7.3.-).
GN ETRL.
OS Brassica oleracea (Cauliflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3712;
PI 1
SEQUENCE FROM N.A.
MEDLINE=98344657; PubMed=9687267;
Chen H.-H., Chang Y.-Y., Yang S.-F., Shaw J.-F.;
"Molecular cloning and sequencing of a broccoli cDNA encoding an ETR-
type ethylene receptor."; PCR98-088.
RL (in) Plant Gene Register; PCR98-088.
CC !- FUNCTION: May act early in the ethylene signal transduction
CC pathway, possibly as an ethylene receptor, or as a regulator of
CC the pathway (By similarity).
CC !- COPACTOR: Binds 1 copper ion per dimer (By similarity).
CC !- SUBUNIT: Homodimer; disulfide-linked (By similarity).
CC !- SUBCELLULAR LOCATION: Endoplasmic reticulum (By similarity).
CC !- PTM: Activation probably requires a transfer of a phosphate group
CC between a His in the transmitter domain and a Asp of the receiver
CC domain (By similarity).
CC !- SIMILARITY: BELONGS TO THE ETHYLENE RECEPTOR FAMILY.
CC !- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
CC !- SIMILARITY: CONTAINS 1 RESPONSE REGULATORY DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF047476; AAC39497.1;
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR003018; GAF.
DR InterPro; IPR003661; His_KinA.
DR InterPro; IPR004359; His_Kin_sig.
DR InterPro; IPR001789; Response_reg.
DR Pfam; PF01590; GAF; 1.
DR Pfam; PF02518; HATPase_C; 1.
DR Pfam; PF00072; response_reg; 1.
DR Pfam; PF00512; signal; 1.

ProDom: PD000039; Response_reg; 1.
DR SMART; SM00065; GAF; 1.
DR SMART; SM00387; HATPase_C; 1.
DR SMART; SM00388; HSKA; 1.
DR SMART; SM00448; REC; 1.
DR PROSITE; PS01019; HIS_KIN; 1.
DR PROSITE; PS01110; RESPONSE REGULATORY; 1.
KW Receptor; Sensory transduction; Phosphorylation; Transmembrane;
KW Transferase; Kinase; Copper; Metal-binding.
FT DOMAIN 350 586 HISTIDINE KINASE.
FT TRANSFAS 609 726 RESPONSE REGULATORY.
FT TRANSMEM 23 43 POTENTIAL.
FT TRANSMEM 54 74 POTENTIAL.
FT TRANSMEM 92 112 POTENTIAL.
FT METAL 65 65 COPPER (BY SIMILARITY).
FT METAL 69 69 COPPER (BY SIMILARITY).
FT DISULFID 4 4 INTERCHAIN (BY SIMILARITY).
FT DISULFID 6 6 INTERCHAIN (BY SIMILARITY).
FT MOD_RES 353 353 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 657 657 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 735 AA; 82240 MW; 520B77291CF2BCC6 CRC64;

Query Match 61.5%; Score 32; DB 1; Length 735;
Best Local Similarity 62.5%; Pred. No. 58;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGHY 10
| : : : :
Db 499 VVPTGSHF 506

RESULT 14
ETRL_ARATH
ID ETRL_ARATH STANDARD; PRT; 738 AA.
AC P49333;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ethylene receptor (EC 2.7.3.-).
GN ETRL OR AT1G66340 OR T27F4.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
PI 1
SEQUENCE FROM N.A. AND MUTANTS ETRL-1; ETRL-2; ETRL-3 AND ETRL-4.
RN MEDLINE=94024007; PubMed=8211181;
Chang C., Kwok S.-F., Bleecker A.B., Meyerowitz E.M.;
"Arabidopsis ethylene-response gene ETRL: similarity of product to
two-component regulators.";
Science 262:539-544(1993).
PI 2
SEQUENCE FROM N.A.
RN STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
Mallitscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
"Sequence and analysis of chromosome 1 of the plant Arabidopsis

```

thaliana.";
 RL Nature 408:816-820(2000).
 [3]
 RN DISULFIDE BONDS, AND MUTAGENESIS OF CYS-4; CYS-6 AND CYS-99.
 RX PubMed-7759498;
 RA Schaller G.E., Ladd A.N., Lanahan M.B., Spanbauer J.M., Bleecker A.B.;
 "The ethylene response mediator ETR1 from Arabidopsis forms a
 disulfide-linked dimer.";
 RT J. Biol. Chem. 270:12526-12530(1995).
 [4]
 RN MUTAGENESIS OF CYS-4; CYS-6; CYS-65 AND CYS-99.
 RP PubMed-8525372;
 RX Schaller G.E., Bleecker A.B.;
 "Ethylene-binding sites generated in yeast expressing the Arabidopsis
 ETR1 gene.";
 RT Science 270:1809-1811(1995).
 [5]
 RN COFACTOR, AND MUTAGENESIS OF GLU-38; CYS-65; HIS-69; HIS-79; MET-87;
 RP MET-104 AND HIS-107.
 RX PubMed-9974395;
 RA Rodriguez F.I., Esch J.J., Hall A.E., Binder B.M., Schaller G.E.,
 Bleecker A.B.;
 "A copper cofactor for the ethylene receptor ETR1 from Arabidopsis.";
 RT Science 283:996-998(1999).
 [6]
 RN SUBCELLULAR LOCATION.
 RP PubMed-11916973;
 RX Chen Y.-F., Randlett M.D., Findell J.L., Schaller G.E.;
 "Localization of the ethylene receptor ETR1 to the endoplasmic
 reticulum of Arabidopsis.";
 RT J. Biol. Chem. 277:19861-19866(2002).
 CC -!- FUNCTION: May act early in the ethylene signal transduction
 pathway, possibly as an ethylene receptor, or as a regulator of
 the pathway.
 CC -!- COFACTOR: Binds 1 copper ion per dimer.
 CC -!- SUBUNIT: Homodimer; disulfide-linked.
 CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum.
 CC -!- TISSUE SPECIFICITY: Leaves, roots, stems, flowers and seedlings.
 CC -!- PM: Activation probably requires a transfer of a phosphate group
 between a His in the transmitter domain and a Asp of the receiver
 domain (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE ETHYLENE RECEPTOR FAMILY.
 CC -!- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 RESPONSE REGULATORY DOMAIN.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; L24119; AAA70047.1; ..
 EMBL; AC020665; AAG52169.1; ..
 InterPro; IPR003594; ATPbind_ATPase.
 InterPro; IPR003018; GAF.
 InterPro; IPR004359; HIS_KIN_sig.
 InterPro; IPR003661; His_kinA.
 InterPro; IPR001789; Response_reg.
 Pfam; PF00072; response_reg; 1.
 Pfam; PF00512; signal; 1.
 Pfam; PF01590; GAF; 1.
 Pfam; PF02518; HATPase_c; 1.
 ProDom; PD000039; response_reg; 1.
 SMART; SM00065; GAF; 1.
 SMART; SM00387; HATPase_c; 1.
 SMART; SM00388; His_kinA.
 SMART; SM00448; REC; 1.
 PROSITE; PS0109; HIS_KIN; 1.
 PROSITE; PS0110; RESPONSE REGULATORY; 1.
 Receptor; Sensory transduction; Phosphorylation; Transmembrane;
 Transferase; Kinase; Copper; Metal-binding.

FT DOMAIN 350 585
 FT RESPONSE REGULATORY.
 FT TRANSMEM 611 729
 FT TRANSMEM 23 43
 FT TRANSMEM 53 73
 FT TRANSMEM 92 112
 FT METAL 65 65
 FT METAL 69 69
 FT DISULFID 4 4
 FT DISULFID 6 6
 FT MOD_RES 353 353
 FT MOD_RES 659 659
 FT MUTAGEN 4 4
 FT MUTAGEN 6 6
 FT MUTAGEN 31 31
 FT MUTAGEN 38 38
 FT MUTAGEN 62 62
 FT MUTAGEN 65 65
 FT MUTAGEN 69 69
 FT MUTAGEN 79 79
 FT MUTAGEN 99 99
 FT MUTAGEN 102 102
 FT MUTAGEN 104 104
 FT MUTAGEN 107 107
 SQ SEQUENCE 738 AA; 82565 MW; D6ED3C4BBA87E96E CRC64;
 Query Match 61.5%; Score 32; DB 1; Length 738;
 Best Local Similarity 62.5%; Pred. No. 58;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 3 VVPXGHHY 10
 DB 498 VVPTGSHF 505
 RESULT 15
 CENC_MOUSE STANDARD; PRT; 906 AA.
 ID CENC_MOUSE STANDARD; PRT; 906 AA.
 AC P49452;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Centromere protein C (CENP-C) (Centromere autoantigen C).
 GN CENPC.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Swiss;
 RX MEDLINE=95048391; PubMed=7959789;
 RA McKay S., Thomson E., Cooke H.;
 "Sequence homologies and linkage group conservation of the human and
 mouse Cenpc genes.";
 RT mouse Cenpc genes.";
 RL Genomics 22:36-40(1994).
 CC -!- FUNCTION: COMPONENT OF THE INNER KINETOCORE PLATE. REQUIRED FOR
 NORMAL KINETOCORE ASSEMBLY.
 CC -!- SUBCELLULAR LOCATION: Nuclear.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; U03113; AAC04314.1; ..

DR MGD: MGI:99700: Cenpc.
 KW Chromosomal protein; Nuclear protein; DNA-binding; Phosphorylation;
 KW Centromere.
 SQ SEQUENCE 906 AA; 102225 MW; 43D529B202E9D71E CRC64;
 Query Match 61.5%; Score 32; DB 1; Length 906;
 Best Local Similarity 62.5%; Pred. No. 72;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 4 VPXGXHYS 11
 |||||
 Db 881 VPSGNHYN 888

Search completed: June 13, 2003, 14:11:46
 Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2003, 14:10:03 ; Search time 78 Seconds
(without alignments)
29.058 Million cell updates/sec

Title: US-09-909-164-42
Perfect score: 52
Sequence: 1 EEVVPXGHVS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*
1: sp._archaea.*
2: sp._bacteria.*
3: sp._fungi.*
4: sp._human.*
5: sp._invertebrate.*
6: sp._mammal.*
7: sp._mhc.*
8: sp._organelle.*
9: sp._phage.*
10: sp._plant.*
11: sp._protist.*
12: sp._virus.*
13: sp._vertebrate.*
14: sp._unclassified.*
15: sp._rivirus.*
16: sp._bacteriap.*
17: sp._archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	69.2	208	2 Q46486	Q46486 corynebacte
2	35	67.3	233	10 Q40479	Q40479 nicotiana t
3	35	67.3	237	10 Q9LW50	Q9LW50 nicotiana s
4	35	67.3	285	16 Q98HU6	Q98HU6 rhizobium l
5	35	67.3	308	16 Q9X2E2	Q9X2E2 thermotoga
6	35	67.3	317	9 Q38317	Q38317 lactobacill
7	35	67.3	360	17 Q27679	Q27679 methanobact
8	35	67.3	678	12 Q9E1X6	Q9E1X6 cercoptihc
9	35	67.3	1028	16 Q8YJ11	Q8YJ11 bruceella me
10	35	67.3	1063	16 Q8RG86	Q8RG86 fusobacteri
11	34	65.4	279	16 Q9RXN9	Q9RXN9 deinococcus
12	34	65.4	322	17 Q9HLH8	Q9HLH8 thermoplasm
13	34	65.4	495	5 Q16912	Q16912 caenorhabdi
14	34	65.4	534	17 Q29966	Q29966 archaeoglob
15	34	65.4	822	2 Q9ZHG7	Q9ZHG7 streptococc
16	34	65.4	825	16 Q99XV4	Q99XV4 streptococc

ALIGNMENTS

RESULT 1

Q46486 PRELIMINARY; PRT; 208 AA.
AC Q46486;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE Hypothetical 23.0 kDa protein (GcrA).
GN GCR.
OS Corynebacterium xerosis, and
OS Corynebacterium striatum.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;
OC Corynebacterium.
OX NCBI_TaxID=1725, 43770;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=C.xerosis; STRAIN=M82B;
RX MEDLINE=96117603; PubMed=8559800;
RA Tauch A., Kassing F., Kallinowski J., Puhler A.;
RT "The Corynebacterium xerosis composite transposon Tn5432 consists of two identical insertion sequences, designated ISI249, flanking the erythromycin resistance gene ermCX.";
RL Plasmid 34:119-131(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=C.striatum; STRAIN=M82B;
RX MEDLINE=20194806; PubMed=10732668;
RA Tauch A., Krief S., Kallinowski J., Puhler A.;
RT "The 51,409-bp R-plasmid pTP10 from the multiresistant clinical isolate Corynebacterium striatum M82B is composed of DNA segments initially identified in soil bacteria and in plant, animal, and human pathogens.";
RL Mol. Gen. Genet. 263:1-11(2000).
DR EMBL; U121300; AAC95478.1; -;
DR EMBL; AF024666; AAG03390.1; -;
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 208 AA; 23012 MW; F1504BE1ECDE85A6 CRC64;

17 34 65.4 1057 16 Q99UR5
18 34 65.4 1442 17 Q96YH5
19 33 63.5 78 6 Q9XST4
20 33 63.5 139 2 Q57489
21 33 63.5 156 16 Q9PC35
22 33 63.5 217 4 Q0404
23 33 63.5 252 16 Q8YWP1
24 33 63.5 299 4 Q9UEE9
25 33 63.5 412 5 Q08523
26 33 63.5 480 5 Q95NK1
27 33 63.5 510 10 Q9SA71
28 33 63.5 722 12 Q9QTC7
29 33 63.5 748 12 Q9YR01
30 33 63.5 795 5 Q9V8K6
31 33 63.5 802 16 Q97QM8
32 33 63.5 816 2 Q9AHT9
33 33 63.5 819 2 Q9ANY3
34 33 63.5 819 10 Q9AVK4
35 33 63.5 819 16 Q97QM9
36 33 63.5 839 16 Q9ANY2
37 33 63.5 844 2 Q9AG74
38 33 63.5 848 16 Q8RG64
39 33 63.5 1039 16 Q9ANV1
40 32 61.5 132 2 Q9SLN2
41 32 61.5 135 4 Q9UBD2
42 32 61.5 143 17 Q8TX62
43 32 61.5 162 16 Q53667
44 32 61.5 201 16 Q9L2F0
45 32 61.5 225 10 Q40129

Q99ur5 staphylococ
Q96yh5 sulfolobus
Q9xst4 canis famil
Q57489 bacteroides
Q9pc35 xylella fas
Q0404 homo sapien
Q8ywp1 anabaena sp
Q9uee9 homo sapien
Q08523 ascaris suu
Q95nk1 plasmodium
Q9sa71 arabidopsis
Q9qtc7 marek's dis
Q9yrl1 ranid herpe
Q9v8k6 drosophila
Q97qm8 streptococc
Q9aht9 streptococc
Q9any3 streptococc
Q9avk4 plsum sativ
Q97qm9 streptococc
Q9any2 streptococc
Q9ag74 streptococc
Q8rg64 thermonaer
Q9ay1 streptococc
Q9sln2 streptomyce
Q9ubd2 homo sapien
Q8tx62 methanopyru
Q53667 mycobacteri
Q9l2f0 streptomyce
Q40129 lycopersico

Query Match 69.2%; Score 36; DB 2; Length 208;
 Best Local Similarity 50.0%; Pred. No. 8.6;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGXHY 11
 Db 130 DVIEPKHYA 139
 :||| | | |

RESULT 2
 Q40479 PRELIMINARY; PRT; 233 AA.
 ID Q40479;
 AC Q40479;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE EREP-2.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxID=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BY4; TISSUE-LEAF;
 RX MEDLINE=95276459; PubMed=7756828;
 RA Ohme-Takagi M., Shinshi H.;
 RT "Ethylene-inducible DNA binding proteins that interact with an
 ethylene responsive element.";
 RL Plant Cell 7:173-182(1995).
 DR EMBL; D38126; BAA07324.1; -
 DR HSP; O80337; 2GCC.
 DR TRANSFAC; T02654; -
 DR InterPro; IPR001471; TF_ERF.
 DR Pfam; PF00847; AP2-domain; 1.
 DR PRINTS; PR00367; ETHRSPLEMT.
 DR ProDom; PD001423; TF-AP2; 1.
 DR SMART; SM00380; AP2; 1.
 SQ SEQUENCE 233 AA; 25563 MW; 6CD16783582COCB5 CRC64;

Query Match 67.3%; Score 35; DB 10; Length 233;
 Best Local Similarity 60.0%; Pred. No. 16;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EVVPXGXHY 10
 Db 90 QAVVPKGRHY 99
 :||| | | |

RESULT 3
 Q9LW50 PRELIMINARY; PRT; 237 AA.
 ID Q9LW50;
 AC Q9LW50;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Ethylene-responsive element binding factor.
 GN NSERF2.
 OS Nicotiana glauca (Wood tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxID=4096;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20399450; PubMed=10945353;
 RA Kitajima S., Koyama T., Ohme-Takagi M., Shinshi H., Sato F.;
 RT "Characterization of gene expression of NSERFs, transcription factors
 of basic PR genes from Nicotiana glauca.";
 RL Plant Cell Physiol. 41:817-824(2000).
 DR EMBL; AB016264; BAA97122.1; -
 DR HSP; O80337; 2GCC.
 DR InterPro; IPR001471; TF_ERF.

DR Pfam; PF00847; AP2-domain; 1.
 DR PRINTS; PR00367; ETHRSPLEMT.
 DR ProDom; PD001423; TF-AP2; 1.
 DR SMART; SM00380; AP2; 1.
 SQ SEQUENCE 237 AA; 26243 MW; 01BC3EEB51E46298 CRC64;

Query Match 67.3%; Score 35; DB 10; Length 237;
 Best Local Similarity 60.0%; Pred. No. 16;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EVVPXGXHY 10
 Db 94 QAVVPKGRHY 103
 :||| | | |

RESULT 4
 Q98HU6 PRELIMINARY; PRT; 285 AA.
 ID Q98HU6;
 AC Q98HU6;
 DT 01-OCT-2001 (TREMBlrel. 18, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Cytochrome c1.
 GN MLL2705.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
 Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti.";
 RL DNA Res. 7:331-338(2000).
 DR EMBL; AP003000; BAB49770.1; -
 DR InterPro; IPR000345; CytC_heme_bind.
 DR InterPro; IPR002326; Cyt_C1.
 DR Pfam; PF02167; Cytochrome_C1; 1.
 DR PRINTS; PR00603; CYTOCHROME_C1.
 DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 285 AA; 30961 MW; 31D9CDE2711747EE CRC64;

Query Match 67.3%; Score 35; DB 16; Length 285;
 Best Local Similarity 55.6%; Pred. No. 20;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 11
 Db 194 VIEGTHIN 202
 :||| | | |

RESULT 5
 Q9X2E2 PRELIMINARY; PRT; 308 AA.
 ID Q9X2E2;
 AC Q9X2E2;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE FTSH protease activity modulator HFLK.
 GN TM1822.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 OX NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MSB8 / DSM 3109;

```

RX MEDLINE-99287316; PubMed-10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linner K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 398:323-329(1999).
RL EMBL; AE001819; AAD36885.1; -.
DR TIGR; TM1822; -.
DR InterPro; IPR001107; Band_7.
DR InterPro; IPR003130; GED.
DR InterPro; IPR001972; Stomat.
DR Pfam; PF01145; Band_7; 1.
DR PRINTS; PR00721; STOMATIN.
DR SMART; SM00302; GED; 1.
DR SMART; SM00244; PHB; 1.
KW Protease; Complete proteome.
SQ SEQUENCE 308 AA; 34778 MW; ADE03603E5101A9D CRC64;

Query Match 67.3%; Score 35; DB 16; Length 308;
Best Local Similarity 75.0%; Pred. No. 22;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
   ||| ||
Db 41 VVPSGIHY 48

RESULT 6
Q38317 PRELIMINARY; PRT; 317 AA.
AC Q38317;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Lysin.
GN Lys.
OS Lactobacillus bacteriophage phi adh.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=12417;
RN [1]
RP SEQUENCE FROM N.A.
RA Altermann E.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-99384014; PubMed-10452953;
RA Altermann E., Klein J., Henrich B.;
RT "Primary structure and features of the genome of the Lactobacillus
RT gasseri temperate bacteriophage phi-adh.";
RL Gene 236:333-346(1999).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE-95138034; PubMed-7836307;
RA Henrich B., Binischofer B., Blaesi U.;
RT "Primary structure and functional analysis of the lysis genes of
RT Lactobacillus gasseri bacteriophage phi-adh.";
RL J. Bacteriol. 177:723-732(1995).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE-93231538; PubMed-8472961;
RA Fremaux C., De Antoni G., Raya R., Klaenhammer T.;
RT "Genetic organization and sequence of the region encoding integrative
RT functions from Lactobacillus gasseri temperate bacteriophage phi-
RT adh.";
RL Gene 126:61-66(1993).
RN [5]
RP SEQUENCE FROM N.A.
RA Engel G., Altermann E., Klein J., Henrich B.;
RT "Structure of a genome region of the Lactobacillus gasseri temperate

```

```

RT phage phi adh covering a repressor gene and cognate promoters.";
RL Gene 210:67-70(1998).
DR EMBL; AJ131519; CAB52540.1; -.
DR InterPro; IPR002053; GH_25.
DR InterPro; IPR003646; SH3_bac.
DR Pfam; PF01183; Glyco_hydro_25; 1.
DR ProDom; PD004620; GH_25; 1.
DR SMART; SM00287; SH3b; 1.
SQ SEQUENCE 317 AA; 34703 MW; 9FF2715EE43561C7 CRC64;

Query Match 67.3%; Score 35; DB 9; Length 317;
Best Local Similarity 66.7%; Pred. No. 23;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 11
   ||| |||
Db 60 VVPMGYHYA 68

RESULT 7
Q27679 PRELIMINARY; PRT; 360 AA.
AC Q27679;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Cell division protein.
GN MTH1642.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DELTA H;
RX MEDLINE-98037514; PubMed-9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AE000923; AAB86115.1; -.
DR InterPro; IPR005140; eRF1_1.
DR InterPro; IPR005141; eRF1_2.
DR InterPro; IPR005142; eRF1_3.
DR InterPro; IPR004405; Pella.
DR Pfam; PF03463; eRF1_1; 1.
DR Pfam; PF03464; eRF1_2; 1.
DR Pfam; PF03465; eRF1_3; 1.
DR TIGRFAMS; TIGR00111; pella; 1.
KW Cell division; Complete proteome.
SQ SEQUENCE 360 AA; 40814 MW; 2A000CB4B3CEFA69 CRC64;

Query Match 67.3%; Score 35; DB 17; Length 360;
Best Local Similarity 45.5%; Pred. No. 26;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGXHY 11
   |::||| |::
Db 98 EDLVPMSHHT 108

RESULT 8
Q9E1X6 PRELIMINARY; PRT; 678 AA.
AC Q9E1X6;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)

```

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Hypothetical 75.9 kDa protein.
 OS Cercopithecine herpesvirus 7.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicellovirus.
 OX NCBI_TaxID=35245;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Gray W.L., Starnes H.B., White M.W., Ashburn C.V., Mahalingam R.;
 RT "Complete Sequence of the Simian Varicella Virus Genome.";
 RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF275348; AAG27217.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 678 AA; 75850 MW; A17B09E30512FE3C CRC64;

Query Match 67.3%; Score 35; DB 12; Length 678;
 Best Local Similarity 50.0%; Pred. No. 52;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 EEVVPXGXHY 10
 II::II I I
 Db 147 EEIIPKTRY 156

RESULT 9

ID Q8YJ11 PRELIMINARY; PRT; 1028 AA.
 AC Q8YJ11;
 DT 01-NAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE ATP-dependent DNA helicase.
 GN BMEI0275.
 OS Brucella melitensis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Brucellaceae; Brucella.
 OX NCBI_TaxID=29459;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;
 RX MEDLINE=20020109; PubMed=11756688;
 RA DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mujer C., Los T.,
 Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
 Jablonski L., Larsen L., D'Souza M., Bernal A., Mazur M., Goltzman E.,
 Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
 Haselkorn R., Kyrpides N., Overbeek R.;
 RT "The genome sequence of the facultative intracellular pathogen
 Brucella melitensis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
 DR EMBL; AE009470; AAL51457.1; -
 DR InterPro; IPR002086; Aldehyde_dehydr.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR001650; Helicase_C.
 DR Pfam; PF00271; helicase_C; 1.
 DR SMART; SM00490; HELIC_C; 1.
 DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 1028 AA; 112996 MW; A752B7042572E219 CRC64;

Query Match 67.3%; Score 35; DB 16; Length 1028;
 Best Local Similarity 54.5%; Pred. No. 81;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 EEVVPXGXHY 11
 I::II I I
 Db 76 EEIVPPGARYS 86

RESULT 10

Q8RG86 PRELIMINARY; PRT; 1063 AA.
 ID Q8RG86
 AC Q8RG86;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)

DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5).
 GN FN0422.
 OC Fusobacterium nucleatum (subsp. nucleatum).
 OC Bacteria; Fusobacteria; Fusobacterium.
 OX NCBI_TaxID=76856;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 25586;
 RX MEDLINE=21886394; PubMed=11889109;
 RA Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
 Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
 Vasieva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
 Larsen L., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
 Fonstein M., Kyrpides N., Overbeek R.;
 RT "Genome sequence and analysis of the oral bacterium Fusobacterium
 nucleatum strain ATCC 25586.";
 RL J. Bacteriol. 184:2005-2018(2002).
 DR EMBL; AE010554; AAL94625.1; -
 KW Ligase; Complete proteome.
 SQ SEQUENCE 1063 AA; 118008 MW; 39700E10B7CCE411 CRC64;

Query Match 67.3%; Score 35; DB 16; Length 1063;
 Best Local Similarity 60.0%; Pred. No. 84;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 2 EVVPXGXHY 11
 I::II I::II
 Db 195 EIVPNGLYNS 204

RESULT 11

Q9RXN9 PRELIMINARY; PRT; 279 AA.
 ID Q9RXN9
 AC Q9RXN9;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Hypothetical protein DR0271.
 GN DR0271.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
 OC Deinococcaceae; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=R1;
 RX MEDLINE=20036896; PubMed=10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 Dodson R.J., Hart D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
 Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium Deinococcus
 radiodurans R1.";
 RL Science 286:1571-1577(1999).
 DR EMBL; AE001889; AAF09867.1; -
 DR TIGR; DR0271; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 279 AA; 31140 MW; DCEA100E0AEE8831 CRC64;

Query Match 65.4%; Score 34; DB 16; Length 279;
 Best Local Similarity 75.0%; Pred. No. 32;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 4 VPXGXHY 11
 I I I I I
 Db 100 VPLGRHYS 107


```
RESULT 12
Q9HLH8
ID Q9HLH8 PRELIMINARY; PRT; 322 AA.
AC Q9HLH8;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE Glucose-fructose oxidoreductase related protein.
GN TA0250.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmatata; Thermoplasmatatales;
OC Thermoplasmatataceae; Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RW Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum."
RL Nature 407:508-513(2000).
DR EMBL; AL445063; CAC11395.1; -.
DR InterPro; IPR000683; GFO_IDH_MocA.
DR Pfam; PF01408; GFO_IDH_MocA; 1.
KW Complete proteome.
SQ SEQUENCE 322 AA; 36918 MW; B8C239EF71009D167 CRC64;

Query Match 65.4%; Score 34; DB 17; Length 322;
Best Local Similarity 75.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGHY 10
DB 66 VVPDGLHY 73

RESULT 13
O16912 PRELIMINARY; PRT; 495 AA.
ID O16912;
AC O16912;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE F10D2.3 protein.
GN F10D2.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Favello A., Fulton L.,
RA Craxton M., Dear S., Du Z., Durbin R., Hillier L., Jier M., Johnston L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Laister N., Latreille P.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., McMurray A., Rooker A., Saunders D., Shownkeen R.,
RA Smalton N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Graves T., Woldmann P., Gilling B.;
RT "The sequence of C. elegans cosmid F10D2."
RT Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
```

```
DR EMBL: AF022972; AAC48234.1; -.
DR InterPro: IPR004151; Sre.
DR Pfam: PF03125; Sre; 1.
SQ SEQUENCE 495 AA; 58190 MW; 0C61139C138EEB4C CRC64;

Query Match 65.4%; Score 34; DB 5; Length 495;
Best Local Similarity 50.0%; Pred. No. 61;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGHY 10
DB 218 ENIVPTGKH 227

RESULT 14
O29966 PRELIMINARY; PRT; 534 AA.
ID AC O29966;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Sarcosine oxidase, subunit alpha (SOXA).
GN AF0273.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -!- COFACTOR: FAD (BY SIMILARITY).
CC EMBL; AE001086; AAB90959.1; -.
DR TIGR; AF0273;
DR InterPro; IPR000564; 2Fe2S_ferredoxin.
DR InterPro; IPR001327; FAD_pyr_redox.
DR InterPro; IPR001041; Ferredoxin.
DR InterPro; IPR000205; NAD_binding.
DR Pfam; PF00111; fer2; 1.
DR Pfam; PF00070; pyr_redox; 1.
DR PROSITE; PS00197; 2Fe2S_FERREDOXIN; 1.
KW FAD; Flavoprotein; Hypothetical protein; Iron-sulfur; Oxidoreductase;
KW Complete proteome.
SQ SEQUENCE 534 AA; 58352 MW; A9DB03174F95093F CRC64;

Query Match 65.4%; Score 34; DB 17; Length 534;
Best Local Similarity 45.5%; Pred. No. 66;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGHYS 11
DB 119 DKVFPHGSHYT 129

RESULT 15
Q9ZHG7 PRELIMINARY; PRT; 822 AA.
ID Q9ZHG7
AC Q9ZHG7;
```

DT 01-MAY-1999 (TReMBLrel. 10, Created)
 DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE Hypothetical 92.4 kDa protein.
 OS Streptococcus agalactiae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Streptococcaceae; Streptococcus.
 OX NCBI_TaxID=1311;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=R268;
 RX MEDLINE=99115568; PubMed=9916102;
 RA Spellerberg B., Rozdzinski E., Martin S., Weber-Heynenmann J.,
 RA Schmitzler N., Luetticken R., Podbielski A.;
 RT "Lmb, a protein with similarities to the Lrai adhesin family, mediates
 RT attachment of Streptococcus agalactiae to human laminin.";
 RL Infect. Immun. 67:871-878(1999).
 DR EMBL: AF062533; AAD13797.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 822 AA; 92386 MW; 80E4EDF313481F98 CRC64;

Query Match 65.4%; Score 34; DB 2; Length 822;
 Best Local Similarity 75.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
 ||| |||
 Db 350 VVPHGDHY 357

Search completed: June 13, 2003, 14:13:59
 Job time : 81 secs